Image-Based Detection Of Neuro-Facial Differences In Foetal Alcohol Spectrum Disorders

Michael Francis John Suttie

Great Ormond Street Institute Of Child Health

University College London

This thesis is submitted for the degree of Doctor of Philosophy

DECLARATION

I, Michael Francis John Suttie confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ACKNOWLEDGEMENTS

Many thanks to my supervisors Prof. Peter Hammond and Prof. Sebastian Ourselin for their support throughout this project. I wish to express my deepest gratitude to Peter for his incredible and ongoing mentoring throughout the early phases of my career, for encouraging me to pursue this PhD, and for the help and support throughout a particularly challenging part of my life.

I also want to thank my wife Rekha for always encouraging, supporting me, and for carrying the little guy who I'm excited to meet.

I dedicate this thesis to my mum and dad, who are sadly no longer with us.

ABSTRACT

Prenatal exposure to alcohol remains as one of the leading, yet preventable, causes of birth defects and neurodevelopmental disorders in the Western world. Over the past 50 years, since the first documented report on the impact of *in utero* alcohol exposure, a broad spectrum of associated effects have been recognised. Foetal alcohol spectrum disorders is the collective term encompassing a range of diagnostic classifications that can be identified. At the most severe end of this spectrum are foetal alcohol syndrome (FAS), recognisable by a characteristic set of facial features, growth delay, neurocognitive deficit, and behavioural impairments. Criteria for either of these diagnostic categories typically requires at least two 'cardinal' facial features: short palpebral fissure length; thin upper lip-vermillion; and, a smooth philtrum. Methods for identifying these features typically rely on subjective observation. This subjectivity means that accuracy of diagnosis is reliant on the skill and experience of the clinician. However, the main clinical challenges arise when an individual presents with confirmed or suspected prenatal alcohol exposure, but without the facial criteria required for FAS diagnoses. These individuals make up the vast majority of the affected population, and clinical recognition can be extremely challenging.

Identification and recognition of facial features associated with prenatal alcohol exposure remain a key area of study. This thesis establishes a novel perspective on the issue of subjective clinical assessment and recognition using 3D face and brain shape analysis. We utilise data from 3D facial imaging, MRI brain images and neurocognitive measures to assess subtle facial differences, face-brain associations and the relationships between face, brain and cognition. Development of innovative techniques and methodologies have allowed us to develop a set of analysis tools applicable to craniofacial assessment, and potentially contribute to the analysis of other facially affected conditions in both clinical and research environments.

CONTENTS

List of Tables

List of Figures

LIST OF ABBREVIATIONS AND ACRONYMS

3D - Three-dimensional
2D - Two-dimensional
AA – Absolute alcohol
ADD – Attention Deficit Disorder
ADHD - Attention Deficit Hyperactivity Disorder
ARBD - Alcohol-related Birth Defects
ARND - Alcohol-related Neurodevelopmental Disorder
CBCL - Child Behaviour Checklist
CC – Corpus Callosum
Cape – Cape Coloured
Cauc – Caucasian
CDSM – Combined Dense Surface Modelling
CIFASD – Collaborative Initiative on Foetal Alcohol Spectrum Disorders
CM - Closest Mean
CN – Caudate Nucleus
CNS - Central Nervous System
CVLT - California Verbal Learning Test
DAS-2 – Differential Ability Scales, Second Edition
DSM – Dense Surface Modelling
FAS – Foetal Alcohol Syndrome
FASD – Foetal Alcohol Spectrum Disorders
FMRI – Functional Magnetic Resonance Imaging
FSIQ – Full Scale Intelligence quotient
GCA – General Cognitive Ability
HC – Healthy Control
HE – Heavily Exposed to Alcohol (but not clinically diagnosable as FAS or PFAS)
IQ – Intelligence quotient
LDA - Linear Discriminant Analysis
MRI - Magnetic Resonance Imaging
NHS – National Health Service
OFC - Occipitofrontal Circumference
PAE – Prenatal Alcohol Exposed
PC – Principal Component
PCA – Principal Component Analysis
PDM – Point Distribution Model
PFAS – Partial Foetal Alcohol Syndrome
PFL – Palpebral Fissure Length
PGI – Philtrum Groove Index
ROI – Region of Interest
SVM - Support Vector Machines
TPS – Thin-plate Spline
UCT - University of Cape Town
WISC-IV - Wechsler Intelligence Scale for Children, 4th Edition

LIST OF APPENDICES

1 INTRODUCTION

Maternal alcohol consumption during pregnancy can have deleterious effects on foetal development. Alcohol's teratogenic insult can be severe at crucial developmental stages. It predominantly damages the central nervous system (CNS) adversely affecting growth, behaviour and cognitive development in childhood when affected individuals are typically brought to the attention of care professionals. Amongst many other detriments, a child prenatally exposed to alcohol can be permanently affected with impaired learning, dysfunctional behaviour and growth deficits. The detrimental impact of *in-utero* alcohol exposure has long been known, with the first formal report documented in the late 1960's (Lemoine *et al.,* 1968). However, this article was in French and did not have a significant impact in medical literature, and there was little interest until 5 years later when the Lancet published a report written by Dr Kenneth Jones & Dr David Smith (Jones and Smith, 1973). Jones and Smith identified a set of clinical diagnostic criteria, in particular they described a specific set of facial characteristics associated with prenatal alcohol exposure (PAE). Since this publication, a spectrum of associated PAE effects has been recognised and has given rise to the term foetal alcohol spectrum disorders (FASDs). There is considerable variation in the combination and severity of effects, convoluting clinical diagnoses, and making the task of identifying those affected clinically challenging. A variety of factors is thought to influence this range and severity, even though the cause predominantly remains the same. In particular, alcohol intake levels, the timing of exposure, exposure patterns, genetic susceptibility and nutrition appear to be key influences on the outcome for individuals with PAE.

This thesis focuses on improving the ability to recognise the facial dysmorphism in FASDs across multiple ethnicities. FASD is a global problem, with recent estimates suggesting the prevalence worldwide to be in the order of 1-2% among the general population (Roozen et al. 2016). At the most severe end of the spectrum are foetal alcohol syndrome (FAS) and partial foetal alcohol syndrome (PFAS); characterised by a set of distinctive facial features, cognitive and behavioural deficits and reduced growth. Three cardinal facial features form part of the FAS/PFAS diagnosis; shortened palpebral fissure length (PFL; eye-opening), thin lip-vermillion border and a smooth philtrum (Figure 1.1). Several diagnostic systems are in circulation which differ on criteria to clinically categorise the effects of PAE, but there is universal agreement amongst these systems

on the significance of the three facial anomalies (Coles *et al.*, 2016). Additionally, across diagnostic guidelines, there is often a necessary combination of criteria for FAS/PFAS, such as a small occipital frontal circumference (OFC; head circumference), a history of maternal drinking during pregnancy, structural brain anomalies, growth deficiency and behavioural or cognitive deficits. Phenotypic facial features are a hallmark of FASDs and play a crucial role in the identification of individuals with PAE. However, it is often the case that a patient may present to a clinician with cognitive and/or behavioural deficits, confirmed or suspected PAE, but lack the identifiable facial criteria for a FAS/PFAS diagnosis. Such individuals may have been exposed to alcohol prenatally at levels similar to those diagnosed as FAS/PFAS and present a vast array of other physical and neurological deficits. Most diagnostic guidelines have some methodology for identifying these individuals. However, the heterogeneous and indistinctive nature of presentation means they are often misdiagnosed or undiagnosed. Furthermore, and most concerning, is that the vast majority of individuals with PAE fall into this clinically challenging category.

Figure 1.1 Image extract from (Wattendorf and Muenke, 2005), four children of different ethnicities with foetal alcohol syndrome presenting the characteristic facial features of narrow palpebral fissures, thin upper lip and smooth philtrum. (A) European descent. (B) Native American (C) African (D) Mixed ethnic backgrounds (Caucasian, African).

Perhaps the most significant effect of PAE is the potential damage caused to the developing brain. Some brain regions are particularly vulnerable, and few remain unharmed by alcohol's teratogenic insult. Magnetic resonance imaging (MRI) enables highly accurate assessment of brain shape, size, spatial displacement and even function. Numerous studies have identified the range of structural and functional brain abnormalities that arise from PAE. Studies consistently report reduced white and grey matter with particular effects on the size and shape alterations

of the corpus callosum (Donald *et al.*, 2015). The corpus callosum, the most extensive white matter tract, is primarily responsible for the interhemispheric connection between the left and right sides of the brain. Significant variations in this structure have been observed, and include complete or partial agenesis (Jones and Smith, 1973), localised shape differences (Bookstein *et al.*, 2007), displacement (Sowell *et al.*, 2001) and localised volumetric reductions (Roussotte, 2012). These findings are representative of the agreed vulnerability of midline structures in PAE. Deep grey matter structures are also consistently reported to be disproportionally affected, and findings include reduced volume of the hippocampi (Archibald *et al.*, 2001; Astley *et al.*, 2009) and basal ganglia (Mattson, Riley, Sowell, *et al.*, 1996; Sowell, 2002) with many studies focusing specifically on the caudate nucleus (Astley *et al.*, 2009; Willford *et al.*, 2010; Joseph *et al.*, 2014; O'Dwyer *et al.*, 2016). The neural structures affected by PAE are responsible for memory, sensorimotor, behaviour and impulse control among many other cognitive functions.

Many individuals diagnosed with FASD exhibit cognitive and behavioural deficits that prevent them from leading productive, independent lives. Typically, they require special attention during their school years and often require continuing support thereafter. Across the FASD spectrum, there is an extreme variability in cognitive dysfunction, although some common functional deficits are apparent in both FAS and in less severe impairments (Mattson, Schoenfeld and Riley, 2001). For example, there are significant declines in overall cognitive ability determined by intellectual quotient (IQ) scores, memory and recall, and executive function (higher level cognitive functions such as abstract thinking, problem-solving, planning ahead). Additionally, psychosocial deficits and problem behaviours such as hyperactivity, impulsivity, poor socialisation and poor communication skills are apparent at a young age, interfering with a child's education, home life and social environments. A combination of these factors put affected children and adolescents at a higher risk of delinquent behaviours, trouble with the law, psychiatric disorders and drug and alcohol abuse, which inevitably follow into adulthood (Streissguth *et al.*, 1991).

As with many childhood conditions, early and accurate diagnosis of FASDs is crucial to introducing appropriate educational, psychological and welfare interventions. Providing the necessary support as early as possible, reduces the risk of social and behavioural problems negatively affecting the individual later on in life. Clinical diagnosis requires a multidimensional approach, taking into account minor physical anomalies, facial dysmorphology and neurocognitive and behavioural profiles. There is a distinct lack of clinical expertise specifically for FASDs, and an individual will often need to seek referral to a specialist clinic where they will typically undergo examination by both dysmorphologists and clinical psychologists. Only a small part of the clinical assessment involves facial analysis, and due to the nature of FASD facial features being mostly observational the examination is necessarily subjective, with the accuracy being dependent on the skill and experience of the clinician.

Craniofacial measurement is used to identify dysmorphic features in affected individuals, typically using manual anthropometry and two-dimensional (2D) photography. For an objective and accurate assessment of features such as philtrum smoothness, thin lip-vermillion or more subtle localised differences in soft tissue we require more robust methods. Identifying the full extent of facial dysmorphism across the FASD spectrum is a primary focus of this thesis as it is a key contributor for diagnosis. We achieve this by applying shape analysis to three-dimensional (3D) face and brain images acquired from stereo-photogrammetric cameras and regional MRI segmentations. 3D facial images provide tens of thousands of points and represent facial

morphology more accurately and in a more convenient form for shape analysis than 2D. Shape analysis of 3D images can take several forms, the most simplistic of which is based on anthropometric measurement derived from landmarks placed on the image surface. 3D analysis of surface shape of the kind advocated in this project supports more sophisticated evaluation and detection of even subtle face shape differences.

The technique at the heart of this project, dense surface models (DSMs) (Hutton, 2004), has been used extensively to analyse 3D facial characteristics associated with neurodevelopmental and related human conditions (Hammond et al. 2005; Hammond et al. 2008; Chinthapalli et al. 2012; Hammond & Suttie 2012; Suttie et al. 2013; Suttie et al. 2017) and face-brain shape in animal models (Lipinski et al. 2012; Heulens et al. 2013). These models have an advantage over landmark-based analysis in their potential to assess shape variation across an entire 3D surface. The DSM algorithm builds surface models from raw 3D data, initially aligning and warping surfaces guided by a series of manually placed landmarks. This provides a dense correspondence, matching points on different face surfaces to produce a shape based on a principal component analysis (PCA) of variation of point displacement from the mean face. Individual 3D face surfaces are then resynthesized as a weighted linear sum of principal components (PCs) that account for the shape variation.

1.1 Aims

The research described here began as the application of DSM techniques and associated software to address a challenging problem: the detection of the facial effects of PAE. Initially, the techniques and software proved effective in performing as well as clinicians in recognising the key facial characteristics FAS, the phenotype towards the more extreme end of FASDs. This thesis describes extensions to the techniques and associated software tools that proved necessary to address the following:

- *Those individuals with confirmed PAE who were not clinically diagnosable with FAS*

  Clinical recognition of the typical FAS face phenotype is possible by trained dysmorphologists, with the assistance of measurement apparatus and photographic guides. However, we seek to improve the identification of the vast majority of individuals affected by PAE presenting with facial effects that are not detectable even by experienced clinicians.

- *The association between the facial and cognitive phenotypes of PAE*

  There is a plethora of complex cognitive impairments associated with PAE, and from our work and others, a facial profile beyond that of the features outlined in clinical guidelines can be identified. We seek to find associations between cognitive and facial presentation in PAE.

- *Potential differences in the facial effects of PAE in diverse populations*

  The high prevalence of FASDs is a global concern. Facial effects in one population may not translate to the same effects in another. Therefore, criteria for the identification of a PAE facial profile may not yield the same results in distinct ethnic groups. We aim to determine if different racial populations present the

typical and non-typical FASD facial phenotypes, ultimately for an improved ability to clinically recognise individuals with PAE.

- *The association between the facial and neural phenotypes of PAE*

  Many studies have elucidated the facial and neurostructural effects of PAE. However, few have addressed the relationship between face and brain effects simultaneously. Developmentally, face and brain are intertwined throughout early gestation when alcohol's teratogenic effects are most damaging. Ultimately, the aim is to utilise the external presentation of the face as a window to the brain.

- *The translation of the methods and software to the FASD clinic to assist diagnosis*

  The overall goal is to provide a screening tool for clinicians for the recognition of facial features associated with PAE. From our previous and continuing work, we can derive a set of proven and optimised tools for the clinical assessment of an individual. Integrating a fine-tuned set of tools to the clinical workflow will provide objective and accurate recognition of facial features to assist the clinician.

1.2 Thesis Structure

Chapter 2 provides a literature review detailing the global issue of FASD, prevalence and definitions of diagnostic subcategories. Additionally, we review the impact on the face and brain of individuals with PAE with an in-depth review of current clinical screening methods.

Chapter 3 demonstrates how DSM based techniques can identify PAE related facial features in a South African cohort, with emphasis on those without clinically recognisable features.

Chapter 4 compares facial effects of PAE two ethnic populations and introduces facial curvature as a new method of delineating features of individuals with PAE.

Chapter 5 introduces an extension to DSM shape analysis for combined shape modelling, allowing the multimodal representation of two 3D surfaces which we apply to face and brain.

Chapter 6 demonstrates progress in translating the methods and analysis described in the previous chapters into a screening tool for clinical use.

Chapter 7 brings together all the main points of the previous chapters and highlights potential future developments and applications of the techniques presented.

The research reported here has benefited from the author's role as a research assistant in the Collaborative Initiative on Foetal Alcohol Spectrum Disorders (CIFASD) programme, a multi-disciplinary study funded by the US National Institute for Alcoholism and Alcohol Addiction (NIAAA). Across several project phases, CIFASD has recruited thousands of subjects acquiring 3D facial images, MRI brain scans, dysmorphology assessments, and data from batteries of neuropsychological and cognitive tests. For this project, the CIFASD programme has provided access to multiple cohorts of children of different age ranges and ethnic backgrounds.

2 BACKGROUND

This chapter explores current and previous approaches taken for the identification of Foetal Alcohol Spectrum Disorders.

2.1 Foetal Alcohol Spectrum Disorders

Foetal alcohol spectrum disorder (FASD) is an umbrella term that encompasses the range of deficits that arise from the teratogenic effects of *in utero* alcohol exposure. At the most severe end of the spectrum are foetal alcohol syndrome (FAS) and partial foetal alcohol syndrome (PFAS); characterised by distinctive facial dysmorphism, neurocognitive deficits and growth delay. The diagnosis of FAS and PFAS strongly relies on the identification of specific facial features typically during a thorough dysmorphology exam undertaken by specialist clinicians. While facial features can often identify individuals at the severe end of the FASD spectrum, a significant proportion of alcohol-exposed individuals do not present such evident phenotypic traits. Heavily-exposed (HE) individuals who have suffered significant in-utero exposure to alcohol, can present without the necessary facial criteria for a FAS or PFAS diagnosis. Depending on the diagnostic guidelines used for clinical assessment, classification without facial features will rely on other criteria requiring confirmed alcohol exposure, central nervous system (CNS) impairments and neurocognitive and behavioural deficits (Chudley *et al.*, 2005; Hoyme *et al.*, 2016). Hoyme et al. 2016 provides an updated set of clinical diagnostic guidelines for diagnosing FASDs, including two categories for confirmed PAE without facial features: alcohol-related neurodevelopmental disorder (ARND), which requires neurobehavioural impairment, and alcohol-related birth defects (ARBD) which requires one or more major malformations. This is just one example of a number of diagnostic guidelines with criteria for the identification of individuals with PAE without facial features (Coles *et al.*, 2016).

Several sets of diagnostic guidelines are in circulation, but discrepancies remain for diagnosing FASD categories (Chudley *et al.*, 2005; Astley, 2006; Farag and Terrace, 2014; Coles *et al.*, 2016). Three notable sets of diagnostic for FASD are: Institute of Medicine (IOM revised); Hoyme Guidelines (Hoyme *et al.*, 2005, 2016); The 4-Digit Diagnostic Code (Astley and Clarren, 1996); and the Canadian Guidelines (Chudley *et al.*, 2005). Since the first publication highlighting the FAS phenotype over 40 years ago (Jones and Smith, 1973) there remains to be ambiguity in the criteria across FASDs. This is primarily due to the heterogeneous presentation of these disorders and the complexity of obtaining metrics for physical features and associated neurocognitive and behavioural profiles. There is, however, universal agreement for four cardinal criteria related to PAE: 1) evidence of maternal drinking during pregnancy 2) restricted growth 3) identifiable physical features (primarily but not limited to the face) 4) neurocognitive and behavioural deficits.

FASDs remain a global issue with high prevalence rates in developed countries of 1-2% of the general population (Popova *et al.*, 2017), and unfortunately, many affected individuals go undiagnosed. This is due to multiple factors, including the lack of facial dysmorphism fitting with diagnostic criteria, unknown maternal alcohol consumption during pregnancy, or the high rate of concurrent mental health issues with FASD presentation (O'Connor and Paley, 2009). Underdiagnosis and misdiagnosis of FASDs are particular issues among children in foster care or in the adoption system. Chasnoff et al. 2015 assessed 547 children in such care, who had at some point been comprehensively examined by a specified mental health clinic - specialising in the assessment and treatment of high-risk individuals. Referral to this clinic was not based on specified screening criteria, but behavioural problems were the most common reason for attendance. Typically, around 30% of referrals to this centre resulted in a FASD diagnosis although they are not regarded as a specific 'FASD' clinic. Within this sample, they identified 156 children who met FASD criteria, of which 125 (80.1%) had not previously been diagnosed as such. For the remaining 31 children who had previously been identified as being affected by PAE, 10 diagnoses were incorrect and changed accordingly, giving a misdiagnosis rate of 6.4%. This study confirmed high rates of PAE in foster care or within the adoption system. Even more concerning is the missed diagnosis and misdiagnosis combined rate of 86.5% for the sample.

2.1.1 Maternal Alcohol Consumption

There are a significant range of PAE related clinical phenotypes which vary in severity and outcome. This variation is thought to be largely dependent on alcohol intake levels, drinking patterns and timing of gestational alcohol consumption (Sulik *et al.*, 1986a; Chudley *et al.*, 2005; O'Leary-Moore *et al.*, 2011; Lipinski *et al.*, 2012; Parnell *et al.*, 2014). Numerous studies have proven alcohol intake considered as 'moderate' to be unsafe during pregnancy (mean absolute alcohol (AA) per week 100ml or 10 units of alcohol (UK)) and substantial enough to have negative impacts on growth, development and neurocognitive outcome during childhood (Day and Richardson, 2004; Dodge *et al.*, 2009; Fraser *et al.*, 2012; Carter *et al.*, 2016). Determining precise drinking patterns and levels in a human population is typically challenging. One study that obtained drinking pattern information, Jacobson et al. (2008), found that binge drinking at least four times per month predicted a reduction in birth weight, at 6.5 and 12 months, slower information processing, and at 13 months poorer psychomotor and mental development. Another study of 500 7-year old children (Bailey et al., 2004) showed children prenatally exposed

to binge patterns of 5 standard drinks (7 UK units of alcohol) had verbal IQ deficits and were 2 ½ times more likely to exhibit clinically significant delinquent behaviours.

Advantageously, animal models can determine the effects of precise ethanol dose and timing of exposure. Lipinski et al. (2012) investigated the timing of alcohol insult at early stages of development in a well-established FASD mouse model. This study used manually segmented high-resolution MRI images to examine correlations between dysmorphology of the brain and face that result from alcohol exposure at specific time points of prenatal development. Time points for ethanol exposure were set to gestational days (GD) 7 and 8.5. Brain components expressed significantly different volumetric measures when comparing GD7 and GD8.5 mice. Some volumetric measures observed in brain components saw a reverse effect when compared to controls. Specifically, GD7 ethanol introduction resulted in decreased olfactory bulb and septal region volumes, whereas the GD8.5 insult resulted in an increased volume in these regions when compared to controls. Morphological differences between face and brain components also showed both similar and opposing shape differences. Overall, GD7 exposed mice produced face and brain correlations that are more consistent with that of the FAS phenotype.

2.1.2 Prevalence

The incidence of FASDs worldwide is challenging to estimate, and epidemiological measures of prevalence vary significantly in the literature. Specific factors influence these estimates, including the method of diagnostic criteria, the epidemiological approach taken and the difficulty of diagnosing FASDs (May and Gossage, 2001). Additionally, there is significant variation in subpopulations of worldwide with higher prevalence rates often correlating with specific demographics of lower socioeconomic status, where access to health care, education and employment are restricted (May *et al.*, 2008, 2010). The most recent estimates suggest that prevalence of individuals with FASDs worldwide to be 1-2% of the general population, and the FAS to FASD prevalence ratio is believed to be around 1:9-1:10 (Roozen *et al.*, 2016). This is a strong indicator that FAS is only a small perceptible share of affected individuals (Chudley, 2008). A comprehensive meta-analysis of 328 studies investigating the prevalence of alcohol consumption during pregnancy, and 62 studies for FAS prevalence, estimated the global prevalence for alcohol use during pregnancy to be in the region of 10% and the FAS prevalence to be around 1.5% (14.6 per 10,000 people) (Popova *et al.*, 2017). For alcohol use during pregnancy, the top five countries were estimated to be Russia (36.5%), UK (41.3%), Denmark (45.8%), Belarus (46.6%) and Ireland (60.4%). The study also found that the top five countries for FAS prevalence per 10,000 people were Belarus (69.1), Italy (82.1), Ireland (89.7), Croatia (115.2) and South Africa (585.3). Notably, the South African FAS rate is higher than that of other countries and vastly higher in relation to alcohol use during pregnancy. The analysis reports this is due to extremely high alcohol consumption per capita, suggesting drinking typically being episodic, binge-like patterns which pose the highest risk of FAS. The South African FAS prevalence rate may have been skewed by the Cape Coloured population and rural communities in the Western Cape where risk factors are at their highest (May *et al.*, 2008). These regions in South Africa have the highest rate of FAS prevalence in the world, as is the consensus with many other studies (Croxford and Viljoen, 1999; May *et al.*, 2000, 2005, 2015; Roozen *et al.*, 2016).

2.2 The Face

2.2.1 Identifying Facial Anomalies Associated with Prenatal Alcohol Exposure

FASDs are characterised by the presence of a multitude of developmental defects. The specific diagnosis of FAS is commonly characterised facially by a smooth philtrum, microcephaly, thin vermillion border, and reduced palpebral fissure length. A number of other minor facial anomalies are also recognised as commonplace in FAS, but are not required for diagnosis (Figure 2.1). To satisfy the criteria for a FAS diagnosis, the patient must present neurodevelopmental deficiencies, delayed growth and at least two of three cardinal facial features (PFL ≤10th percentile, smooth philtrum and thin upper lip) (Jones *et al.*, 2006; Moore *et al.*, 2007; Hoyme *et al.*, 2016). Such features are identified in a clinical examination undertaken by a trained dysmorphologist, but only a few can be measured on a continuous scale. PFL is typically measured using a plastic ruler held up to the face and tilted to account for the angle of the zygoma (Figure 2.2c), or by estimating it from manually place points on a 2D image (Astley & Clarren 2001). Techniques used to identify lip and philtrum features often employ a visual assessment using an ethnically specific photographic lip-philtrum scoring chart (Astley and Clarren, 1996, 2000; Hoyme *et al.*, 2016) as shown in Figure 2.3. Inevitably, such charts are categorical at best and their interpretation, when set against the face of a child, is subjective.

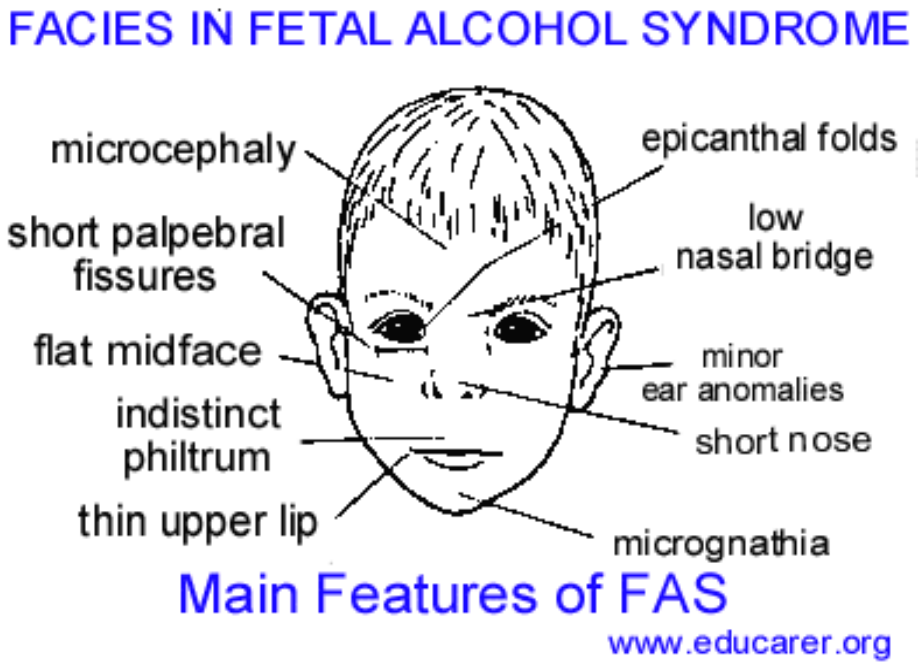
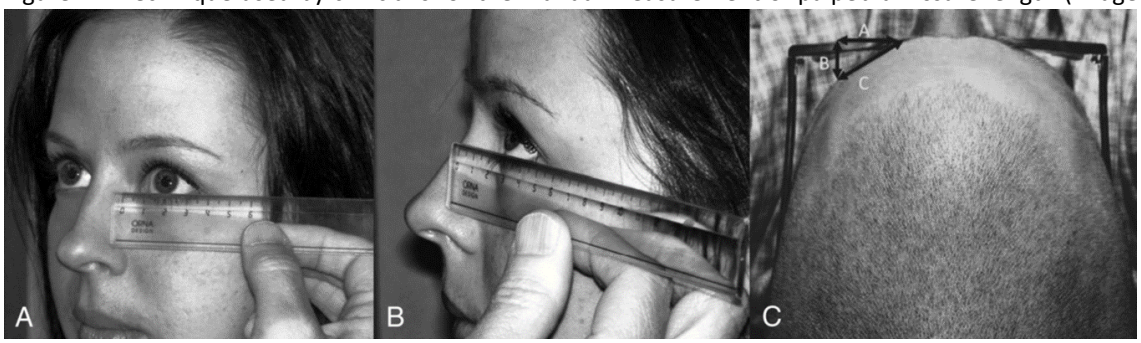


Figure 2.1 Full range of facial features associated with FAS (source: www.educarer.org)

Figure 2.2 Technique used by clinicians for the manual measurement of palpebral fissure length (image



extract from Hoyme et al. 2016). A small plastic ruler is used to measure the distance between endocanthion and exocanthion points for both eyes for which the average is taken. (A) With the clinician sitting at an equal height, the subject is asked to look up so that the ruler can be positioned as close to the eye as possible. (B) The ruler can be rested on the cheeks for stability and is required to be angled slightly to correct for the curve of the zygoma. (C) The angle of the ruler corrects for distance "B" (angle AC).

Figure 2.3 Astley and Clarren Lip-Philtrum Guide (image extract from Astley & Clarren 2000)(A) Philtral smoothness and thin upper lip-vermillion measured on a scale of 1 to 5 (1 = least affected, 5 = most extreme). The subject must have a neutral facial expression as any tension on the lip area will alter the resulting score. Clinicians select the guide that best matches the patient's ancestral origin, in this example either (B) Caucasian or (C) African.

2.2.2 Clinical Recognition of Prenatal Alcohol Exposure Associated Facial Features using 2D

Commercial off the shelf software for the identification of the FAS facial phenotype is currently available. Initially developed in 2003, the FAS Facial Photographic Analysis Software uses 2D photographs in combination with the 4-Digit Diagnostic Code (Astley and Clarren, 1996, 2000, 2001) which aims to identify the magnitude of expression of the FAS phenotype. The software determines metrics to represent the three key facial features; PFL, lip thickness and philtrum smoothness from a series of three 2D images using manually guided procedures. The image acquisition protocol requires three images; a frontal view, lateral view and ¾ view to be taken with a standard digital camera. For dimensionality, the image is scaled using a visual guide (small sticker) physically placed on the subject's forehead. Using the three images uploaded to a computer, the user can then measure or estimate the three cardinal facial features.

Firstly, the PFL is calculated from the front portrait by manually annotating points to represent exocanthi and endocanthi. PFL is calculated as the mean of left and right distances which is weighted using an adjustment factor, subsequently adding 7% to the distance. The weighting constant is representative of the angle of the zygoma, shown as distance "B" in Figure 2.2c, to account for the loss of perspective in 2D. The mean left-right PFL measure taken using this method, with the adjustment factor applied, was shown by the authors to have just a 1% margin of error compared to subjects measures with a calliper (Astley and Clarren, 2001; Astley, 2015). However, the use of a 2D photographic assessment of PFL have has come under some scrutiny by other researchers and is not always the preferred method by clinicians (Cranston *et al.*, 2009; Hoyme *et al.*, 2016). This lack of agreement is primarily due to potential inaccuracies caused by individual variation of the angle of the zygoma (Figure 2.2c), questioning the ability to account for this correction using a single constant value. Studies testing the concordance of techniques to record PFL have found measurements taken using 2D imaging to be consistently shorter than those measured using face-to-face techniques (Cranston *et al.*, 2009; Avner *et al.*, 2014). It is for this reason that normative data used to determine percentile rank for an individual should be consistently derived from 2D images.

The software also offers a semi-automated tool for calculating thinness of the upper lip vermillion by a measure of circularity. Circularity is calculated as the perimeter of the upper lip squared divided by the enclosed area (perimiter^2/area). This requires the user to manually trace an outline of the upper lip on the front image using the cursor (Figure 2.4). Thinner, FAS-like upper lips are indicated by a greater circularity. Values for circularity are then matched relative to a position on the lip-philtrum guide (Figure 2.3). This method, whilst continuous is intended to subsequently assist the clinician with their rank selection on the lip-philtrum guides.

Figure 2.4 Extract from Astley & Clarren 2001. Upper lip circularity requires a manually traced region delineated by the vermillion border. The area within this perimeter is calculated and circularity is calculated as the perimiter²/area.

Philtrum smoothness is determined in much the same way as with a live assessment using a 5-point lip-philtrum guide (Figure 2.3), but instead using the ¾ and frontal images. However, as was described in an early case definition study (Astley and Clarren, 1996), philtrum smoothness was calculated as a measure of luminosity. Firstly, a contour was drawn horizontally across the philtrum, between the upper lip and subnasale, equal in length to the distance between the left and right cheilion points (corners of the mouth) and a width 1/5th of the upper lip to subnasale distance. Using a 256 greyscale image, the pixel values along this contour are divided into 100 averaged segments of equal size to represent luminosity. To generate a measure of philtrum smoothness, the darkest pixels in the philtrum furrow were subtracted from the brightest at the philtrum ridge. The deeper the philtrum furrow, the greater the contrast of the ridge relative to the furrow. In this earlier study, the authors tested reliability of this measure against a Likert scale, which was used to set a rank for the luminosity values. Although the publication describes this method as providing a "high degree of accuracy and precision" for measuring philtral smoothness, there is no mention of this technique in the latest versions of the software and little mention in the literature of studies preceding Astley & Clarren 1996.

2.2.3 Face Analysis using 3D

2D imaging suffers limitations of perspective and dimensionality, and analysis of surface shape and finer detail are restricted. Several studies have used the highly detailed and accurate reconstructions of the human face provided by 3D photography and statistical analysis techniques to address this problem for the identification of FAS facial features (Meintjes *et al.*, 2002; Mutsvangwa *et al.*, 2010). The earliest of these studies (Meintjes *et al.*, 2002) utilised stereo-photogrammetry systems to obtain reliable 3D images of the face to determine clinically relevant anthropometric measurements. Before this investigation, previous studies using photographic methods for distance measures had been performed only on single 2D images (Astley and Clarren, 1996). As previously mentioned, measurements imperative to diagnosis using 2D photography, such as PFL, have to be synthetically adjusted to account for the points being off the midline perspective. The ability for 3D images to reliably capture points that lie in different planes highlights the inadequacies of single planar 2D photographic techniques.

Obtaining point-to-point distance measures using 3D images has shown to be a reliable method comparable to anthropometric measures attained physically by a trained dysmorphologist (Aynechi *et al.*, 2011).

Moore et al. 2007, utilised 3D images of 276 subjects (aged 2.8 to 21.2 yrs), clinically labelled as FAS or control from 4 different ethnic populations (Cape Coloured (37%), Finnish Caucasian (36%), African American (9%) and North American Caucasian (18%)). The hypothesis of this study was based on the assumption that a unique combination of anthropometric features would best discriminate FAS and controls in each of the four ethnic populations. To test this, images captured using a commercially available laser scanning system were manually annotated with 20 anthropometric landmarks. Discriminant analysis was performed to classify FAS and controls utilising an optimal combination of age at evaluation and 16 measurements derived from the landmarks. From the discriminant analysis, the Finish cohort performed with the highest overall classification rate (93%) and utilised nine anthropometric measurements (inner/outer canthal distance, PFL, midfacial depth, ear length, nasal and nasal bridge length and bitragal width). The discriminant function calculated from the Cape Coloured cohort achieved a classification accuracy of 92% and included five measurements (inner canthi distance, philtrum length, ear length, minimal frontal width and bizygomatic width). For the African American cohort, the classification rate of 79% was achieved using only two measures (PFL and philtrum length). The North American Caucasian population achieved the lowest classification rate of 77% using only the inner and outer canthi distances. This study showed that anthropometric measures derived from 3D images could discriminate FAS from controls across a wide age range and multiple ethnicities. It also highlights and explicates ethnic differences in the presentation of FAS, which appear to be particularly discriminating in the Cape Coloured cohort.

More sophisticated approaches employing statistical shape analysis techniques were later introduced to characterise the facial anomalies associated with FAS. Landmark-based morphometric analysis techniques, such as those defined in Mutsvangwa et al. (2010), demonstrate the ability to detect features present in FAS affected individuals from 3D facial images. By utilising a landmark-based discriminant function analysis model, the authors were able to evaluate the mean shape differences between FAS and control groups at two different time points (5yrs and 12yrs) achieving overall classification accuracies of 95% and 80% respectively. As well as demonstrating the value of 3D morphometric analysis for identifying individuals with FAS, this analysis supports the notion that FAS facial features are more pronounced in younger generations and subsequently diminish with age (Astley and Clarren, 1996).

Another study of facial morphology of prenatally exposed individuals revealed differences in directional asymmetry (Klingenberg *et al.*, 2010). 3D scans of children from a Finnish Caucasian cohort (n=90) and a South African mixed-ancestry cohort (n=78), were obtained and manually annotated with 17 anthropometric landmarks. Using geometric morphometrics methods they analysed shape data from the landmarks to show the degree of directional asymmetry for each individual. The analysis showed average directional asymmetry in the FAS groups to be significantly higher than those in the control groups, primarily consisting of a leftward shift in midline landmarks. These results show the extent of facial dysmorphology in alcohol-exposed individuals goes beyond what is clinically recognisable, and can include subtle differences in asymmetry.

2.2.4 Dense Surface Modelling

While landmark based studies can accurately define geometric morphometry, its shortcoming when applied to FAS is the inability to define surface curvature in soft tissue areas such as the philtrum. The technique at the heart of this project, dense surface modelling (DSM), has been utilised in previous studies to discriminate atypical facial and brain morphology associated with a range of genetic conditions (Hammond *et al.*, 2005, 2008; Hammond and Suttie, 2012; Lipinski *et al.*, 2012). DSM provides a technique for building point distribution models (PDM) (Cootes *et al.*, 1995) of 3D surfaces from raw 3D point data. These models have an advantage over traditional landmark-based analysis in their potential to assess variation across the entire surface in terms of 3D shape rather than linear and angular measures or even co-location of landmarks.

The first stage of computing a DSM requires the placement of a sparse set of anatomical landmarks. Landmarks are utilised to form a dense correspondence of points to be established between all surface meshes. To achieve this, we compute the mean landmarks using the generalised Procrustes algorithm (Goodall, 1991), which serve as a frame of reference to warp each surface to using thin-plate spline warping (TPS) (Bookstein, 1997). Each set of surface landmarks is now in exact alignment, and we bring all surfaces into close alignment using the same smooth transform applied for each surface mesh. The dense correspondence is then computed using a base mesh of points as reference finding the closest point on each surface from each vertex of this mesh. Each surface is now represented by a warped set of resampled points with relative connectively to the base mesh. Final preparation of the surfaces comes in form of 'unwarping' these surfaces to return them to their original form.

Building the PDM follows methodology from Cootes et al., 1995. Corresponding vertices for each surface and can now be treated as landmarks and the mean shape is computed by applying the Procrustes algorithm to these points. Next, we apply a principal component analysis (PCA) to represent the variation of the densely corresponded points. Individual 3D surfaces are resynthesized as a weighted linear sum of principal components (PCs) that account for 99% of the shape variation. The standard Euclidean-based metric determines distance and overall shape difference between vectors of weights synthesising individual faces and average faces of homogeneous groups. Pattern recognition techniques such as closest mean, linear discriminant analysis and support vector machines can then be applied in a supervised learning fashion to classify individuals and/or groups previously categorised clinically. Localised shape differences can be visualized by heat mapping point-to-point distances for example to discriminate differences between average faces of control and affected subgroups.

2.3 Prenatal Alcohol Effects on The Developing Brain

Prenatal alcohol exposure adversely affects the brain from the earliest stages of development. Causing midline anomalies, ranging from minor structural defects to full agenesis of the corpus callosum (Bookstein *et al.*, 2007; Lebel, Roussotte and Sowell, 2011), small brain volume, cerebellum defects (Archibald *et al.*, 2001; Riley, McGee and Sowell, 2004) and volumetric reductions of the basal ganglia (Mattson, Riley, Sowell, *et al.*, 1996). Initial case studies of structural brain damage caused by PAE were only obtainable from autopsies of severely affected children (Jones and Smith, 1973; Clarren and Smith, 1978a). These cases were representative only of those with the most extreme damage, as children do not typically die during infancy as

a result of alcohol teratogenesis. Nevertheless, they were valuable for understanding the damage inflicted on developing structures and revealed malformations of the corpus callosum (including full agenesis), ventricle dilation and smaller cerebellum. The vast majority of cases consistently revealed microcephaly (small head) or microencephaly (small brain) in individuals with FAS/PFAS. In more recent times, advancements in imaging technology have allowed for the in vivo detection of subtle changes in brain morphology across the entire spectrum. Magnetic resonance imaging (MRI) supports functional assessment, linear measurements, volume and shape changes in the more subtle cases across the FASD spectrum.

MRI has been extensively used over the past decade by researchers to quantify structural and functional defects present in individuals with known prenatal alcohol exposure. Recent and continuing advancements in MRI analysis techniques have provided comprehensive neurostructural assessment, consistently showing significant volumetric reductions and abnormalities in several brain components; including the parietal lobe, basal ganglia, cerebellum, hippocampus and cortical regions (Riley, McGee and Sowell, 2004; Klingenberg *et al.*, 2010; Lebel, Roussotte and Sowell, 2011). Previous studies also indicate more subtle anomalies such as atypical patterns of grey matter distribution, structural morphometric differences and structural displacement (Riley, McGee and Sowell, 2004; O'Dwyer *et al.*, 2016). Several other regions of the brain have also been identified as few structures remain unharmed by alcohol's teratogenic insult (Lebel, Roussotte and Sowell, 2011). The full extent of brain anomalies attributable to prenatal alcohol exposure often results in severe neurocognitive impairment and a range of behavioural disorders exhibited in early childhood which continue later in adulthood (Roussotte *et al.*, 2012). Later, in this thesis, we focus on two brain regions; the corpus callosum and the caudate nucleus, both of which exhibit disproportionate sensitivity to in-utero alcohol exposure (Archibald *et al.*, 2001; McGee, 2005; Norman *et al.*, 2009; Joseph *et al.*, 2014).

2.3.1 Corpus Callosum

The corpus callosum is a white matter mid-line structure responsible for the inter-hemispheric connection of the two cerebral hemispheres. Consisting of around 200 million axons, this structure's primary role involves directing motor activity from opposing sides of the body along a small group of fibres called the anterior commissure. Consequently, with the use of diffusion tensor imaging (DTI) techniques, inter-hemispheric functional connectivity in this region has been shown to be significantly lower in children with FASD than in controls (Wozniak *et al.*, 2011). Studies have also shown the corpus callosum to be involved with many neurocognitive tasks, such as memory, verbal learning, thought processing, visuospatial ability and executive function (Sowell *et al.*, 2001; Bookstein *et al.*, 2002). Current research indicates that damage to white matter may be disproportionate in FASDs, and it is certain that the corpus callosum is particularly vulnerable to the teratogenic effects of alcohol (Archibald *et al.*, 2001; Riley, McGee and Sowell, 2004).

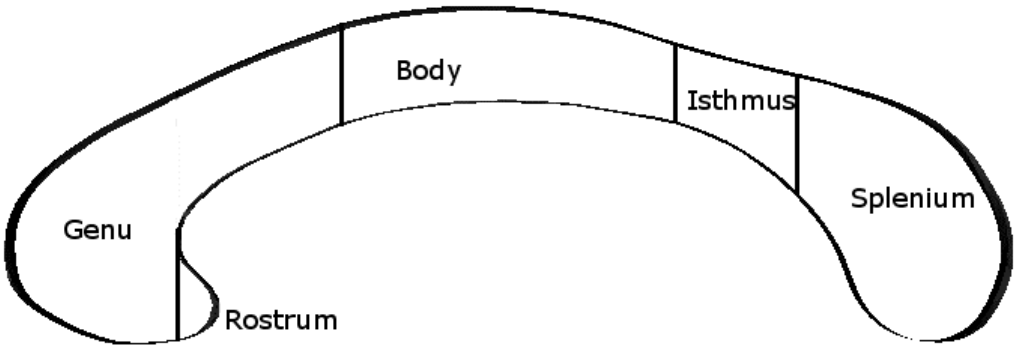


Figure 2.5 Anatomy of the corpus callosum. Mid-sagittal view of the mean control (n=47) extracted from a corpus callosum DSM model and divided into anatomical sections as per (Bookstein *et al.*, 2002). The anterior sections rostrum and genu, the central body, and the anterior, isthmus and splenium.

Corpus callosum formation typically transpires during weeks 12-13 of gestation. From then on, growth in a caudal direction continues up to weeks 18-20. Development starts with the formation of the genu, followed by the body and the selenium. The rostrum, the anterior section which curves underneath, does not develop until the final stages of corpus callosum development (anatomy of the corpus callosum shown in Figure 2.5). By week 20 the structure is fully formed, but it has been suggested that corpus callosum growth does not reach full growth maturity until the 3rd decade of life (Riley *et al.*, 1995a). Given the rapid development of the brain at early stages of life and the clinical necessity for an early diagnosis, it is essential to quantify abnormalities present in alcohol-exposed individuals at the earliest possible time point. Bookstein et al. (2007) observed unique morphological changes in the corpus callosum in early infancy. Structural images of 23 highly alcohol-exposed and 21 unexposed infants from 1.4 to 16.7 weeks after birth were acquired using transfontanelle ultrasonography. This produced a mid-sagittal viewpoint in which corpus callosum shape could be observed. This study identified a localised anomaly described as a “hook” shape between the splenium and the long arch of this plane in individuals with FAS. This was objectively characterised as an angle of fewer than 90 degrees which was present in 12 of the 23 affected individuals and in only 1 of the 21 unexposed controls. The study emphasised the complexity and heterogeneity of structural brain morphology present in affected individuals.

Several studies have identified a variable range of severe abnormalities in alcohol-exposed individuals compared to controls (Riley *et al.*, 1995b; Bookstein *et al.*, 2002, 2007; Lebel, Roussotte and Sowell, 2011; Wozniak *et al.*, 2011). Region-specific callosal thickness differences have been observed in affected subjects showing that specific segments of the corpus callosum may be more severely affected than others (Sowell *et al.*, 2001). The anterior-most section - the genu, and posterior-most sections – the istmus and splenium, are frequently quoted as being the most affected regions (Sowell et al., 2001; Wozniak et al., 2009). Findings have also shown significant correlations between corpus callosum morphology and the functional deficits that result from prenatal alcohol exposure (Sowell *et al.*, 2001; Bookstein *et al.*, 2007; Dodge *et al.*, 2009; Yang, Phillips, *et al.*, 2012; Yang, Roussotte, *et al.*, 2012). Sowell et al. (2001) observed displacement of the splenium and isthmus, which lie more anterior and inferior in alcohol-exposed individuals compared to controls. Displacement was evident across the FASD spectrum

but less so in a prenatally exposed subject who did not have FAS facial criteria. This indicates the presence of neuroanatomical deficits similar to those in FAS can be apparent, but to a lesser degree, in individuals without facial dysmorphology. This study also revealed some brain-behaviour associations, primarily verbal learning deficits which is more correlated with the degree of displacement in the alcohol-exposed group compared to the controls.

More recently, MRI image analysis of the corpus callosum in prenatally exposed individuals has extended to include new-borns (Jacobson *et al.*, 2017). Before this study, MRI imaging had not investigated the feasibility of detecting alcohol-related brain effects in infants. Segmentation of infant brain structures is challenging due to a lower contrast-to-noise ratio, smaller head size and a short scan time which is dependent on the infant's non-sedated sleep patterns. Additionally, adult brain profiles and segmentation atlases do not correspond with infant brain morphology. Hence, this study used manual tracing to extract a cross-sectional area of the corpus callosum for analysis. From a South African Cape Coloured cohort, 32 infants were born to heavy drinking mothers who had consumed ≥14 standard drinks a week or had patterns of binge drinking with 4 or more drinks per occasion. 11 control infants were recruited where the mother had abstained from alcohol or only drunk very minimal quantities during pregnancy. The study found corpus callosa of the heavily exposed individuals to be significantly smaller after controlling for total intracranial volume, than that of the control group. Additionally, they found no association with infant corpus callosum volumes and sex, gestational age, age at the time of scan, smoking, methamphetamine or marijuana use during pregnancy. This study shows that damage to brain structures can be determined at an early age and supports early identification for a range of structural impairments. Recognition of structural brain defects may be an early indicator of future cognitive impairments and early identification could assist in providing the correct form of early-stage intervention and support.

2.3.2 Basal Ganglia

The basal ganglia are a set five subcortical nuclei responsible for a range of processes including sensorimotor coordination, cognitive processing, associative learning, memory and behavioural reinforcement (BOLAM *et al.*, 2000; Sizemore *et al.*, 2016). Located within the forebrain and midbrain, the basal ganglia consists of the putamen, globus, substantia nigra, subthalamic nucleus and the caudate nucleus. This group of nuclei appear to be particularly vulnerable to alcohol teratogenesis and have been the focus of study since the availability of MRI analysis (Riley *et al.*, 1995a; Mattson, Riley, Sowell, *et al.*, 1996; Nardelli *et al.*, 2011; Roussotte, 2012). Notably, an early detailed volumetric analysis found reductions of basal ganglia in exposed individuals compared to controls (Mattson, Riley, Sowell, *et al.*, 1996). Furthermore, when they divided the structure into its individual parts, the caudate nucleus and lenticular nuclei were significantly reduced in the children with FAS compared to controls. When controlling for the overall reduction in intracranial volume the proportional volume of the caudate nucleus remained significantly reduced in the FAS subjects. This is a key indicator of the selective damage exerted by alcohol teratogenesis, and the disproportionate impact it has on the caudate nucleus.

More recent studies have assessed the brain-behaviour relationships that exist in children with a history of prenatal alcohol exposure (Coles *et al.*, 2011; Fryer *et al.*, 2012; Roussotte, 2012). Fryer and colleagues aimed to characterise how structural brain differences in alcohol-exposed subjects related to alcohol-associated cognitive deficits. In this study, 7 unexposed controls and 21 subjects (ages 9–21 yrs) with a confirmed history of heavy prenatal alcohol exposure (defined

as gestational consumption ≥14 standard drinks per week, or at least 1 binge per week of ≥4 drinks), underwent a battery of neurobehavioral tests and a structural brain MRI. Neurocognitive testing focused cognitive control and verbal learning and recall - two key domains with deficits associated with prenatal alcohol exposure. Regional brain volumes were calculated for six key regions of interest (ROI); bilateral frontal, parietal, and temporal lobes, cerebellum, caudate nuclei, and hippocampi. Using regression analyses of ROI volumes and cognitive measures they observed volumetric decrements to be a predictor of cognitive performance. The most consistent, and significant associations were recorded for the bilateral volume of the caudate nucleus all cognitive variables recorded and remained a significant predictor of cognitive performance, even after accounting for confounding variables (total brain volume, IQ, and age). This finding supports the notion that brain-behaviour relationships exist in the particularly vulnerable structures in alcohol-exposed individuals. More specifically, the volume of the caudate appears to be strongly associated with cognitive control and verbal learning and recall.

Volumetric analyses of brain components provide an insight into damage caused by prenatal alcohol exposure. However, an understanding of shape morphology gives an in-depth perspective of structural deficits. Sophisticated 3D surface based techniques allow analysis of shape variation, and in one study have been applied to brain MRI segmentation in alcohol-exposed individuals (Joseph *et al.*, 2014). This analysis uses surface deformation-based techniques to assess shape variation in the caudate nuclei and hippocampi of children with FASD. Images of 19 controls and 12 subjects with FAS and PFAS diagnoses were imaged using high-resolution MRI systems. The authors calculated surface mesh representations from manual region segmentations of each structure, and an iterative closest point algorithm was used to register a template mesh of a control subject across all surfaces in the dataset. The resulting calculation achieves a set corresponding points across all surfaces representing the exact geometry of each component. PDMs (Cootes *et al.*, 1995) were constructed to describe shape variability across the set by quantifying variability in the positions of the corresponding points. Values expressing modes of variation are calculated through principal component analysis (PCA) of the covariance matrix of point coordinates. Localised Hotelling T2 methods were used to analyse shape variance between control and exposed sets. Shape analysis of the caudate nucleus showed significant differences between the FAS/PFAS and control groups, showing changes in the head and tail regions of the structure. The degree of change was also shown to correlate with the level of alcohol exposure. This study demonstrates the potential for utilising 3D surface based techniques to assess the degree of structural deformation in the brain components of alcohol-exposed individuals, providing far more comprehensive detail than volumetric measures alone.

2.3.3 Face-Brain Relationships

Localised shape variation in the corpus callosum has been associated with facial measures, showing a correlation between reduced callosal thickness and a shortened PFL (Yang, Phillips, *et al.*, 2012). Additionally using volumetric measures of several subcortical brain structures incorporating total brain, grey matter, and total white matter volumes, Roussotte et al. (2012) found relationships between volume metrics, facial dysmorphology, and neurocognitive function in FASD. Standard facial scoring metrics; PFL and philtrum-lipometer scores, acquired from the clinical exam, were shown to correlate with volumetric reductions in regional brain

structures. A reduction in PFL was significantly associated with a reduction in the left thalamus, right caudate and left and right ventral diencephalon. Correlations were found between philtrum lipometer scores and the putamen, thalamus and pallidum with higher scores being significantly associated with smaller volumes in the left and right thalamus and in the left pallidum and right putamen. These studies provide evidence for patterns of concurrent insult on face and brain, in particular the effects and relationships of midline structural development.

2.4 Neurocognitive and Behavioural Impairments

An extensive and heterogeneous array of cognitive and behavioural impairments associated with PAE have been identified in the literature (Mattson and Riley, 1998; Roebuck, Mattson and Riley, 1999; Mattson, Schoenfeld and Riley, 2001; McGee, 2005; Roussotte *et al.*, 2012). Initially, studies focused on general cognitive impairment and intellectual quotient scores (IQ), however, these metrics fail to distinguish specific cognitive ability across the whole spectrum (Mattson & Riley 1998). Recent studies of FASD cohorts have revealed consistent impairments of specific neurocognitive and behavioural functions, while other functions are spared (Janzen, Nanson and Block, 1995a; Wozniak *et al.*, 2011, 2016; Lewis *et al.*, 2015a).

Individuals with PAE show particular deficits in language and memory (Janzen, Nanson and Block, 1995b; Lewis *et al.*, 2015b). An early study (Mattson et al. 1996), administered the California Verbal Learning Test - Children's Version (CVLT-C) to assess children with FAS. This test is designed to evaluate a child's immediate and delayed recall and recognition memory using a word list. The study identified individuals with FAS aged 5-16 had difficulty learning and recalling words after a short delay and made a significant number of errors compared to matched controls. Additionally, the FAS children found it challenging to distinguish between target and distractor words with a significantly higher false-positive rate on recognition testing. The study also controlled for mental-age of the FAS participant and found some of the FAS-control differences remained. However, FAS children did perform equally well as controls for information recall tasks. These findings indicate that individuals with FAS have a unique pattern of learning and recognition deficits which coincide with initial 'encoding' phases of memory development. A more recent study (Lewis *et al.*, 2015b) looks further into the effects on encoding, in heavily-exposed populations and those exposed to alcohol in moderate levels. Two populations were studied; African Americans from Detroit, and a South African Cape Coloured cohort. This study identified moderately-exposed children to have significantly lower recognition discrimination scores compared to controls, and delayed word recall scores to be just shy of significant ($p=0.06$). Additionally, the study found all CVLT-C scores to correlate with a measure of prenatal alcohol exposure in the Cape Coloured cohort, and a handful of scores correlated in the Detroit cohort.

Another learning domain focuses on visual-spatial ability, for which deficits are also prominent in individuals with PAE (Kaemingk, 2003). This affects learning ability for tasks that involve spatial relationships among objects. In one study, poor performance was recorded in FAS children when asked to perform delayed object recall, but their immediate recall was in line with controls (Uecker and Nadel, 1996). It was also noted that FAS and control individuals were equal in their ability to recall common household and schoolroom objects. One test in this study involved replacing an object back to its original position on a table, for which FAS participants performed poorly compared to controls.

Perhaps one of the most commonly known and recognisable hallmarks of PAE affected children is the associated attention deficit hyperactivity disorder (ADHD). This aspect of the behavioural phenotype is so closely correlated that PAE affected children are often misdiagnosed as ADHD and consequently receive incorrect treatment. However, Coles et al. 1997 found differences in behavioural characteristics between PAE children and unexposed children with a primary diagnosis of ADHD. Reported individuals with PAE were found to present uniquely different scores on the Child Behaviour Checklist (CBCL) (caregiver report used for the identification of problem behaviours). Specifically, children primarily diagnosed with ADHD had significantly higher T-scores than both PAE and control children scoring higher on CBCL scales for attention problems, delinquency and aggressive behaviour. Coles and colleagues found that PAE children had a greater ability to focus and maintain attention, but struggled more to shift attention from one task to another compared to those diagnosed with ADHD. This study was one of the first to suggest that children with PAE, and those with a primary diagnosis of ADHD, to have disparate and delineating neurocognitive and behavioural characteristics.

Executive function is an umbrella term for the frontal lobe controlled higher level cognitive functions that revolve around intricate everyday tasks such as such as abstract thinking, problem-solving, planning ahead, set shifting, organisation and inhibition (Rasmussen, 2005). Deficits in executive function are commonplace in an individual with FASD, but in particular, across the literature three main domains appear consistently identified; set shifting (ability to unconsciously shift attention between one task and another), working memory and inhibition. PAE children demonstrate difficulties with set shifting if, for example, they are asked to switch from naming objects from one category such as animals, to another, such as plants (Kodituwakku *et al.*, 1995). Also, executive function deficits appear to be expressed in both FAS and individuals with PAE who do not present with physical features. Mattson et al. (1999) assessed 18 children with FAS or PAE without criteria for FAS, as well as a control set. A battery of neurocognitive tests to measure cognitive flexibility, inhibition, planning, word context and verbal reasoning was administered to all participants. Both PAE groups showed significant impairments in all executive function domains relative to controls, and with the exception of the set switching test, there were no significant differences between the two PAE groups. Hence, this study shows that cognitive and executive function deficits are present in individuals with PAE lacking the facial features and criteria for a FAS diagnosis.

3 FACIAL DYSMORPHISM IN FOETAL ALCOHOL SPECTRUM DISORDERS

The three cardinal facial features of Foetal Alcohol Syndrome (FAS) are thin upper lip, smooth philtrum and reduced palpebral fissure length (PFL). Clinical recognition of these features is sometimes challenging, particularly at the mild end of the Foetal Alcohol Spectrum Disorder (FASD) spectrum. In this chapter, we analyse facial dysmorphism in a Cape-Coloured population of South Africa which has a high incidence of prenatal alcohol exposure. We aim to improve the detection of facial dysmorphism across the FASD spectrum, with an emphasis on heavily alcohol-exposed individuals who do not exhibit the classic facial features of FAS.

3.1 Introduction

Our primary aim is to improve the understanding of facial dysmorphism across the FASD spectrum. The ultimate goal is to assist clinicians in the recognition of prenatal alcohol exposure in children with cognitive and behavioural deficits, across all age ranges, to enable adequate social support and behavioural/educational intervention to be put in place. In this chapter, we apply existing dense surface modelling (DSM) techniques (Hutton, 2004; Hammond and Suttie, 2012) to a South African Cape Coloured population where there is a high incidence of prenatal alcohol exposure. We cover the full range of the FASD spectrum from the most severe phenotype to those with prenatal alcohol exposure (PAE) without the characteristic physical features. The main body of results derives from a study (Suttie *et al.*, 2013) completed during the initial phase of this PhD project. Collaborators listed as authors on the paper undertook the data collection and clinical examinations, and the author performed facial analysis under the supervision of the principal investigator Prof Peter Hammond.

The diagnostic categories used in this chapter refer to the alcohol-exposed groups; foetal alcohol syndrome (FAS), partial foetal alcohol syndrome (PFAS) and non-syndromal heavily exposed (HE). At the severe end of the FASD spectrum, FAS and PFAS were diagnosed based on the criteria outlined in Table 3.2 (detailed in section 3.3.2) requiring the identification of facial dysmorphology, neurocognitive deficits, growth and brain defects. The HE classification is used where PAE is confirmed, but clinical requirements for a FAS or PFAS diagnosis are not met. The HE subset remains clinically challenging due to the absence of identifiable characteristic criteria, and their PAE can be overlooked and their diagnosis inaccurate or late. It is estimated that individuals with less readily recognised effects of PAE represent a population ten times that of FAS (May *et al.*, 2000).

Previous studies using DSM based shape analysis have discriminated atypical facial morphology in a variety of conditions (Hammond *et al.*, 2005; Cox-Brinkman *et al.*, 2007; Chinthapalli *et al.*, 2012; Heulens *et al.*, 2013b). DSM models have an advantage over traditional landmark-based analysis in their potential to assess variation across the entire surface, utilising 3D shape rather than linear and angular measures or even co-location of landmarks. Across the FASD spectrum, facial morphology ranges from characteristic, recognisable features to indistinctive and clinically challenging presentations. The typical FAS facial features of reduced PFL, smooth indistinct philtrum and a thin upper lip vermillion overlap with other genetic and teratogenic conditions that clinicians consider to be differential diagnoses (Table 3.1). Previous studies utilising 3D images and statistical analysis techniques have sought to discriminate between control and FAS populations (Meintjes *et al.*, 2002; Moore *et al.*, 2007; Fang *et al.*, 2008; Mutsvangwa *et al.*, 2010). However, few have identified or recognised any distinctive facial dysmorphism in individuals with PAE who fall outside the FAS-PFAS categorisation.

Condition	Overlapping features with FASD	Differentiating features
Aarskog	Small nose with anteverted nares, broad philtrum, maxillary hypoplasia, and wide-spaced eyes	Rounded face, down-slanted palpebral fissures, widow's peak, crease below lower lip, incomplete out folding of upper helices, and dental eruption problems
Cornelia de Lange	Long philtrum, thin vermillion border, anteverted nares, and depressed nasal bridge	Single, bushy eyebrow extending across forehead, long eyelashes, downturned mouth, high arched palate, and short limbs/stature
Dubowitz	Short palpebral fissures, wide spaced eyes, and epicanthal folds	Shallow supraorbital ridge with nasal bridge near the level of the forehead, and broad nasal tip
Foetal hydantoin	Wide-spaced eyes and depressed nasal bridge	Short nose with bowed upper lip
Foetal valproate	Epicanthal folds, anteverted nares, long philtrum with thin vermillion border, and wide-spaced eyes	Short nose with bowed upper lip
Noonan Syndrome	Low nasal bridge, wide-spaced eyes, and epicanthal folds	Down-slanted palpebral fissures, keratoconus, wide mouth, and protruding upper lip
Williams Syndrome	Short palpebral fissures, anteverted nares, long philtrum, depressed nasal bridge, and epicanthal folds	Wide mouth with full lips, stellate pattern of the iris, periorbital fullness, and connective tissue disorders
Maternal phenylketonuria fetal effects	Epicanthal folds, short palpebral fissures, long underdeveloped philtrum, and thin vermillion border	Small upturned nose, round faces, and prominent glabella
Toluene embryopathy	Short palpebral fissures, midface hypoplasia, smooth philtrum, and thin vermillion border	Micrognathia, large anterior fontanel, downturned mouth corners, hair patterning and ear abnormalities, and bifrontal narrowing

Table 3.1 Conditions considered to be differential diagnoses for FASD. Extract from Suttie et al., 2013.

In this chapter, we test FASD clinical categorisation using face shape to discriminate individuals with PAE from healthy controls (HC). Given the facial phenotypes associated with FAS and PFAS, classification based on face shape alone is predictably more concordant with clinical diagnosis than it is for the HE subgroup. The heterogeneity of the facial presentation in the latter is an obstacle for recognition using traditional methods that rely on group-wise similarity. To analyse this heterogeneous group we apply a clustering technique, signature graph analysis (Hammond and Suttie, 2012; Hammond, Suttie, *et al.*, 2012) which clusters individuals according to the similarity of their facial dysmorphism defined in terms of normalised difference from matched controls. Facial signature graphs constructed from exposed individuals revealed facial dysmorphism in half of the HE set to be more FAS-like than control-like with an associated correlation of cognitive impairment.

3.2 Cape Coloured Population

The 'Cape Coloured' population is an ethnic group in South Africa (SA) primarily composed of mixed ancestry individuals. They form the predominant populace of the Western Cape region but are a minority in SA constituting around 9% of the total population. This mainly rural community established themselves in the 17th century through the interracial mixing of European colonisers, indigenous Khoisans, Xhosa people, and slaves imported from the Dutch East Indies. This diverse ancestral heritage has resulted in a racially heterogeneous population, who typically self-identify as 'Coloured'.

Cape Coloured communities typically suffer from diminished socioeconomic status, likely to be the result of the socio-political history of racial inequality in the Republic of South Africa (RSA). During apartheid, access to schooling, healthcare and other public services was restricted based on ethnicity. As a result, the standard of living, income and education have suffered significantly amongst Cape Coloured communities (Myers *et al.*, 2013). Although conditions have improved post-apartheid, Cape Coloured populations remain significantly underrepresented in the use of alcohol and drug treatment facilities. Racial inequality plays less of a part than in the apartheid years, but there remain barriers to the available services with a reduced awareness of therapy options, geographic accessibility and affordability (Myers, 2013). The presence of these barriers is an unwelcomed influence on the patterns of alcohol and drug misuse, which remain to be significantly higher than the national average (May *et al.*, 2008; Phillip Gossage *et al.*, 2014).

Agriculture in RSA is one of the largest employment sectors. Wine accounts for 15% of agricultural production, and The Western Cape Province is the hub for wine production. Job opportunities offered to Cape Coloured communities are predominantly in this sector, in particular women (London, 1999), and historically, workers in farms were given alcohol as an employment benefit. This practice known as the 'dop' system became outlawed in 1961. However, a legal loophole allowed workers to be rewarded with alcohol as a bonus or gratuity (Phillip Gossage *et al.*, 2014). Remnants of this practice have had adverse influences on the attitudes for those who work in wine production and agriculture, where excessive and episodic drinking is the cultural norm. Most concerning is the high frequency of alcohol consumption during pregnancy. In one study assessing alcohol use by pregnant women in a Western Cape Province population (Croxford and Viljoen, 1999), nearly half (42.8%) of expectant mothers admitting to varying degrees of alcohol consumption. Over half of the individuals who drank, did so at harmful levels of consumption to put their unborn children in a 'high risk' category for FAS.

It is evident that population characteristics of the Cape Coloured community contribute to a significantly increased prevalence of FASDs. Standards of education, socioeconomic status and mothers from rural locations are key associated maternal risk factors (May *et al.*, 2000). Even within the community, there is a variance in severity and dose-dependent links with more complex risk factors (May *et al.*, 2008). The given factors associated with this population results in FASD prevalence amongst the highest in the world.

3.3 Methods

3.3.1 Dataset

The dataset used for this study consists of 192 participants from a Cape Coloured community local to the University of Cape Town (UCT). This dataset forms part of 2 UCT longitudinal cohorts (n=137, n=55) with previously published studies (Jacobson et al. 2008, Jacobson et al. 2011). In the larger subset (Jacobson et al. 2008), gestational alcohol consumption information was obtained prospectively at the point of recruitment using a timeline follow back method (Jacobson et al., 2002). This interview process was carried out during pregnancy in antenatal clinics to assess first and second-trimester alcohol intake, then six weeks postnatally to obtain third-trimester intake. Alcohol-exposed individuals with a diagnosis of PFAS (n=26) and HE (n=75) were all subject to maternal consumption levels of at least 7oz absolute alcohol per week (equivalent to a minimum of 21 UK units per week or greater than or equal to 14 standard drink measures) or occasional binges of 5 or more drinks. The FAS (n=22) classification is irrespective of confirmed alcohol exposure, and the control (n=69) set contains only individuals whose mothers reported abstinence during pregnancy. Pregnant women who reported alcohol consumption were referred to specialist services and are advised to lower alcohol intake or abstain. In the second cohort (Meintjes *et al.*, 2010; Jacobson *et al.*, 2011), participants were primarily recruited by screening undertaken in a rural school (n=31) in an area where alcohol abuse is abundant. The 24 remaining were older siblings of those in the first cohort.

3.3.2 Clinical Assessment

Participants were examined independently by expert dysmorphologists; Dr Eugene Hoyme and Dr Luther Robinson, blinded to maternal alcohol exposure information, following Institute of Medicine (IOM) protocol (Hoyme *et al.*, 2005) for FASD diagnoses. Upper lip thinness and philtral smoothness were measured using the Astley and Clarren Lip-Philtrum Guide (Astley and Clarren, 2000). There was a definitive agreement between clinicians on the measurements including palpebral fissure length (PFL) (r=0.77), philtrum (r=0.80) and lip scores (r=0.84). Additionally, another clinician Dr Nathaniel Khaole, who examined a subset was also in agreement with these scores (median r = 0.78). All clinicians were in agreement with the final FASD diagnostic labels assigned. Subjects with reported abstaining mothers were assigned to the HC category unless they met the criteria for FAS. Three children of mothers who reported abstinence from alcohol during pregnancy met the criteria for a FAS diagnosis.

At a nine-year follow-up stage, clinicians administered standardised batteries of neuropsychological tests. These include the Wechsler Intelligence Scale for Children, 4th Edition (WISC-IV; Wechsler 2003) to test verbal ability, functional IQ and perceptual reasoning, and the California Verbal Learning Test – Children's Version (CVLT-C; (Delis et al. 1994) a well-established

verbal memory test. The CVLT-C tests are used to measure word recall for words learned over successive trials and recognition memory after a 20-minute delay. All neuropsychological tests were administered by a Master's level neuropsychologists in the child's primary language; either English or Afrikaans.

Five feature domains exist within this clinical protocol (Hoyme *et al.*, 2005); alcohol exposure, facial dysmorphism, growth deficiencies, structural brain defects and neurocognitive/behavioural deficits. Table 3.2 illustrates how these domains fit diagnostic categorisation. Testing positive for alcohol exposure requires a confirmed alcohol intake during pregnancy (prospectively or retrospectively recorded) of 1oz of absolute alcohol (3 UK units) per day or regular binge drinking sessions involving 5 or more drinks per occasion. Facially, the guidelines consider the three characteristic FAS facial features: shortened PFL (≤10th percentile) measured using a ruler and a thin upper lip and smooth philtrum score of 4 or 5 on the Astley & Clarren guides. Growth deficiency is recorded by taking height and weight measurements at the time of assessment and comparing them to standardised population records corrected for racial norms. To include brain defects requires evidence of abnormal morphogenesis or deficient brain growth, marked positive if the head circumference is ≤10th percentile compared to racially adjusted norms, or the presence of structural brain abnormalities. Behaviourally, a child will test positive if they express a complex pattern of cognitive impairment or behavioural characteristics, not consistent with their age, which cannot be explained by other causes (e.g. family background, genetic, environmental causes). These impairments can include, but are not limited to, learning impairment, poor planning, language disorders, behavioural disorders (ADD/ADHD), motor dysfunction, communication and poor social interaction.

	Criteria	**FAS**	**PFAS**	**HE**
Alcohol exposure	Confirmed excessive gestational alcohol consumption	-	✓	✓
Face	PFL (≤10th percentile) Thin upper lip (Astley scale 4 or 5) Smooth philtrum	≥2 of 3	≥2 of 3	-
Growth	Height weight (≤10th percentile)	✓	≥1 of 3	-
Brain defects	Head circumference (≤10th percentile) Structural abnormality (CNS)	≥1 of 2		-
Behaviour	Behavioural or cognitive deficit	-		-

Table 3.2 Summary of the criteria used for categorising FAS and PFAS, and the delineation from nonsyndromic HE children. FAS is diagnosed with or without exposure requiring 2 or more cardinal facial features, growth deficiency ≤10th percentile and the presence of a brain defect. PFAS requires alcohol exposure confirmation but has a greater inclusion of growth, brain and behavioural associations. The HE category does not necessitate traits other than excessive maternal alcohol intake (minimum 7oz absolute

alcohol per week (≥21 UK units per week or ≥14 standard drink measures) or occasional binges of 5 or more drinks.

3.3.3 Facial Analysis

Imaging Protocol

3D facial images were captured using a static, tripod-mounted commercial 3D camera from 3DMD *(www.3dMd.com).* This high precision stereophotogrammetry system is capable of capturing an 180-degree ear-to-ear image with an order of 25,000 mesh points at sub-millimetre accuracy (geometric resolution <0.2mm). Each participant within this study had their 3D photo taken as part of the clinical assessment. Using in-house software (Facemark), the author manually annotated a sparse set of 24 reliable anatomical landmarks (Gwilliam, Cunningham and Hutton, 2006) illustrated in Figure 3.1 to each 3D surface.

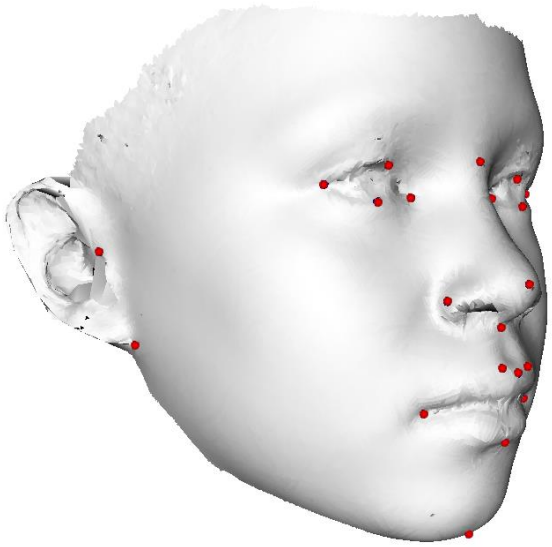


Figure 3.1 Landmark placement used for point correspondence. Left and right: exocanthion, endocanthion, palpebral inferius, palpebral superius, tragion, otobasion inferius, crista philtrum, cheilion, and alare. Mid-line points: Nasion, gnathion, subnasale, pronasale, labiale superius and labiale inferius.

Dense Surface Modelling

DSMs of facial components were built using in-house software (Hammond and Suttie, 2012). DSM applies principal component analysis (PCA) to represent the variation of thousands of densely corresponded points, after surface alignment and thin-plate spline warping using the sparse set of manually placed anatomical landmarks (Figure 3.1). DSM models resynthesize individual 3D surfaces as a weighted linear sum of principal components (PCs) that account for 99% of the shape variation. The standard Euclidean-based metric determines distance and overall shape difference between vectors of weights synthesising individual faces and average faces of homogeneous groups. We applied multi-folded classification techniques; closest mean (CM), linear discriminant analysis (LDA) and support vector machines (SVM) in a supervised learning fashion to classify individuals and groups previously categorised clinically. For example, for closest mean classification, we calculated the relative similarity of a facial surface to the average control or mean FASD subgroup face to discriminate between sets of unseen randomly sampled test examples across 20 cross-validation trials. We determined discrimination between

two subgroups in terms of the mean area under the receiver operating curves (ROC) of the 20 trials, representing the probability of correctly classifying a pair of randomly selected faces with each of the pair from a different subgroup.

Additionally, localised shape differences were visualised by normalised heat mapping of the displacement of surface mesh points (Figure 3.2; Facial Signatures) to discriminate differences between average faces of sex-ethnicity matched controls and alcohol-exposed subgroups.

Facial Signatures and Facial Signature Graphs

Before this project, advances in DSM methodology introduced *face signatures* and *face signature graphs* (Hammond and Suttie, 2012). We previously applied these techniques in a study (Hammond, Suttie, *et al.*, 2012) to identify atypical facial morphology in a rare connective tissue disorder; *Fibrodysplasia ossificans progressiva* (*FOP*). The analysis quantified, for the first time, patterns of facial dysmorphology in FOP individuals using their face signature and associated *face signature graphs.* Facial signature graphs provide an analysis tool for identifying unique subsets of facial dysmorphism and hence have a useful application in FASD given the heterogeneity of facial effects expressed across the spectrum.

Facial signatures discriminate statistically significant regions of facial difference from a matched comparison group. Throughout this thesis, we generate running means from consecutively aged controls (n=35) providing ethnicity-aged-matched control faces as reference. Using 35 surfaces constructs matched-means that proves to be stable when testing the distance from each individual to a series of means generated from randomly sampled control faces (see Appendix 1). In determining the facial signature of an individual care is taken to omit them from the control subset used to determine the matched mean. For each of the ~25,000 points on a face, we calculate mean and standard deviation values of this displacement relative to those of the matched controls. The vector of such normalised displacements across the densely corresponded surface points is then used to derive distance measures of any given subject from another subject or subgroup mean. This can also be visualised as a heat map representing regions of statistically significant shape difference from the norm. Utilising anthropometric landmarks on the mean face of the DSM model, we can define axes (horizontal: between exocanthi; vertical: nasion to gnathion corrected to be perpendicular to the horizontal axis) to calculate face signatures of directional displacement, e.g., lateral, vertical and anterior-posterior.

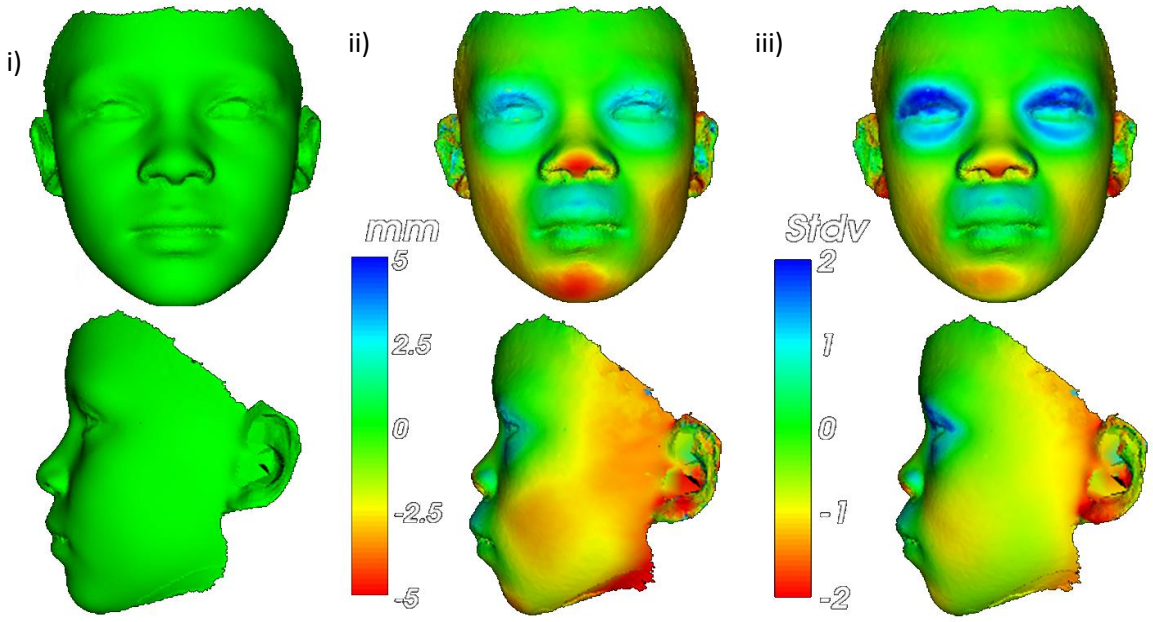


Figure 3.2 i) Mean face of 35 Cape Coloured controls ii) Cape Coloured FAS individual compared to matched mean showing surface-normal displacement expressed as a raw distance measure (±5mm) iii) The same individual with a facial signature heat map overlaid. Regions of colour now represent the significance of dysmorphology; with heat map scaling set to -2 (red) and +2 (blue) standard deviations. This visualisation allows us to identify regions of dysmorphism; malar flattening (red in the mid-face); philtral smoothness (blue on the philtrum centre); retrognathia (red on the chin) and shortening of the nose (red on columella/underside of the nose).

These localised, normalised displacement values support a rudimentary estimate of facial dysmorphism as the square root of the sum of the squared normalised displacement values for each point on a face. This will be high for severely dysmorphic or outlier faces and small for those with little difference from their matched mean.

Visualising facial signature heat maps allows the manual detection of dysmorphism for individuals and can similarly be utilised to assess group mean differences. However, large datasets are challenging to analyse in this manner, so instead, we take an automated approach to identifying similarities and differences in dysmorphology using clustering. Initially, we define a metric expressing the difference between one facial signature and another. For pairs of face signatures, the *face signature distance* (FSD) calculates the Euclidean distance between the representative signature vectors. If the normalising group is a set of healthy controls, then the FSD between two faces, for example, is a straightforward and reasonable estimate of the difference in their facial dysmorphism. FSD is used to calculate the distance between each facial signature with every other in the dataset, constructing a matrix of distances for every individual. For each face, minimum FSD values are retained and used to draw directional edges from one signature to another. The length of an edge between two vertices is the FSD between them with lower values equating to more similar dysmorphism. The shortest FSD of a face to another signifies a link to the individual with the closest dysmorphism. Face signatures connected by an edge form sub-clusters. The resultant clusters are similarly connected by the shortest distance from one cluster to another to produce a fully interlinked signature graph. Facial signature

graphs will typically cluster individuals based on the similarity of dysmorphic features expressed i.e. reflecting face shape differences from the norm and not face shape similarity.

3.4 Results

3.4.1 Sample Characteristics

The male-female split for the 192 subjects was balanced without age distribution differences. Reported gestational alcohol intake did not differ statistically between the exposed groups (FAS, PFAS & HE). Mothers of the exposed group reported an average daily intake of 2.8 standard drinks, with excessive binge patterns on weekends consuming an average of 8.9 drinks per session. Additionally, drug use was reported by the alcohol-consuming mothers with 12 using cannabis, 3 using methaqualone ("mandrax") and one cocaine user. A significant majority (70.8%) of mothers reported smoking during pregnancy, of which 16.7% averaged over ten cigarettes per day. Across the sample, we observed the presence of label-dependent effects for head circumference, FAS < PFAS < HE & HC; for height and weight, FAS < HE & HC; and PFAS < HE & HC, for weight only.

	FAS (n=22)	PFAS (n=26)	HE (n=75)	HC (n=69)	F or χ^2
Age at 3D photo (years)	10.6 ± 2.4	10.0 ± 1.5	10.4 ± 2.7	10.1 ± 2.6	0.46
Sex n (% male)	12 (54.5%)	14 (53.8%)	36 (48.0%)	34 (49.3%)	0.47
Parity	2.9 ± 1.4	2.8 ± 1.9	1.7 ± 1.0	2.0 ± 1.2	8.27***
Height (cm)	127.0 ±	131.0 ± 10.4	136.6 ± 13.6	135.6 ± 14.1	3.68*
Weight (kg)	25.8 ± 7.2	27.8 ± 6.4	34.0 ± 10.9	34.2 ± 13.2	5.29**
Head circumference (cm)	49.9 ± 1.4	51.6 ± 1.4	53.0 ± 1.7	53.0 ± 1.9	22.55***
BMI	15.7 ± 1.6	16.0 ± 1.3	17.8 ± 3.2	18.0 ± 3.9	5.05**
Facial anomalies					
Short palpebral fissures	18 (81.8%)	18 (69.2%)	9 (12.0%)	8 (11.6%)	72.89***
Flat philtrum	20 (90.9%)	25 (96.2%)	22 (29.3%)	15 (21.7%)	69.13***
Thin vermillion	20 (90.9%)	25 (96.2%)	20 (26.7%)	17 (24.6%)	68.34***
Child's WISC IV-IQ	60.6 ± 11.1	65.3 ± 10.4	71.1 ± 12.8	75.3 ± 12.0	7.82***
Alcohol use during	1.8 ± 2.4	1.2 ± 1.2	1.3 ± 1.8	0.0 ± 0.0	14.65***
oz AA/occasion	4.7 ± 2.8	4.2 ± 2.8	4.4 ± 4.0	0.0 ± 0.0	35.33***
Frequency (days/week)	2.0 ± 1.9	1.9 ± 1.0	1.7 ± 1.4	0.0 ± 0.0	36.65***
Cigarettes/day during pregnancy	6.3 ± 6.2	7.4 ± 5.3	8.0 ± 7.1	4.3 ± 7.4	3.68*

Table 3.3 Values represent mean ± SD or % and F statistic calculated using one-way ANOVA analysis. Pearson's chi-squared test ($\chi 2$) applied to sex and for the presence of facial anomalies where the criteria are: short PFL (≤10th percentile); smooth philtrum and thin vermillion border ranked 4 or 5 on Astley Lip-Philtrum Guide. AA=absolute alcohol; 1oz equal to 2 standard drinks (~3 UK units). PFL measures were obtained during the clinical examination using a rigid ruler and compared with published age-matched norms. PFAS and HE mothers smoked more than those of controls (p = 0.05 and 0.1, respectively). For

parity, mothers of FAS and PFAS were higher than those of controls and HE (all p=0.01). FAS children were born to older mothers than those of HE and controls (p=0.01) and PFAS mothers were older than HEs (p=0.05). Dose-dependent relations for head circumference: FAS, PFAS, HE and controls (p < .001); height and weight: FAS, HE and controls (P < .01); weight: PFAS, HE and controls (p < .01). WISC IQ: FAS, HE and controls (p < .002); PFAS, controls (p < .002); PFAS, HE (p = .055); HE, controls (p < .10); FAS and PFAS, controls (p < .005); FAS, HE (p < .005); PFAS, HE (p <.06); HE, controls (p < .10).

* P= 0.05.

** P= 0.01.

*** P= 0.001.

3.4.2 Facial Growth & Dysmorphology

In a face DSM model with a mixed age range, principal component 1 (PC1) will typically reflect facial growth. Predictably, FAS facial growth is significantly reduced compared to HCs using PC1 as a scale (Figure 3.3.i) unattributed to age distribution differences. Individuals with PFAS also show reduced growth, however to a lesser extent (Figure 3.3.i), and HEs show only a negligible growth reduction compared to HCs (Figure 3.3.ii). Facial signature heat maps of mean FAS (Figure 3.3.iii) and PFAS (Figure 3.3.iv) groups are almost monochromatic due to the severity of growth reduction. As growth deficits are part of the FASD criteria, we built DSM models representing both size and shape. When size variance dominates subtle shape differences in this way, we can apply scaling to average faces for comparison to visualise shape only differences. Figure 3.4 shows FAS, PFAS and HE averages scaled using the nasion to gnathion distance as a ratio to match the HC comparison mean.

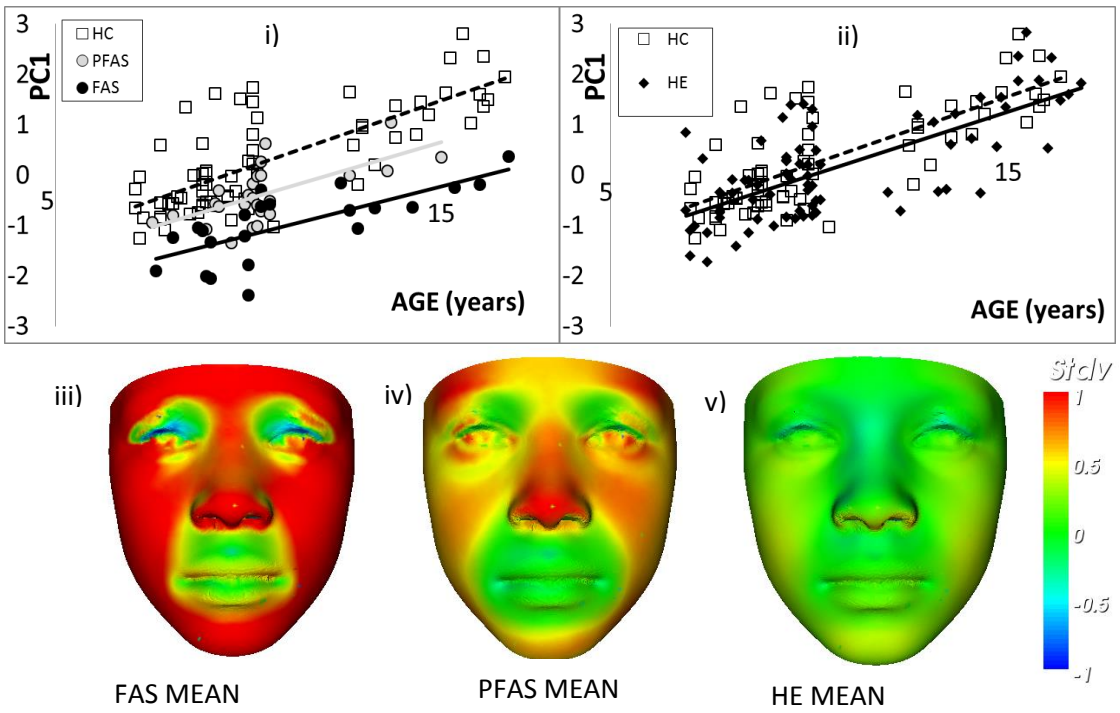


Figure 3.3 Facial growth comparison with controls (HC) using PC1 (i) significantly lower facial growth for both FAS ($p < 0.001$) and PFAS ($p < 0.001$). (ii) HE showing only a marginal reduction ($P = 0.052$). This reduction of facial growth has a substantial influence on average group facial signatures with FAS (iii) and PFAS (iv) showing red-dominant colouring across the majority of the face. Sensitivity to the subtler facial features are diminished as a result. (v) The HE mean is not affected in the same degree as FAS and PFAS means at the same SD (±1) scale.

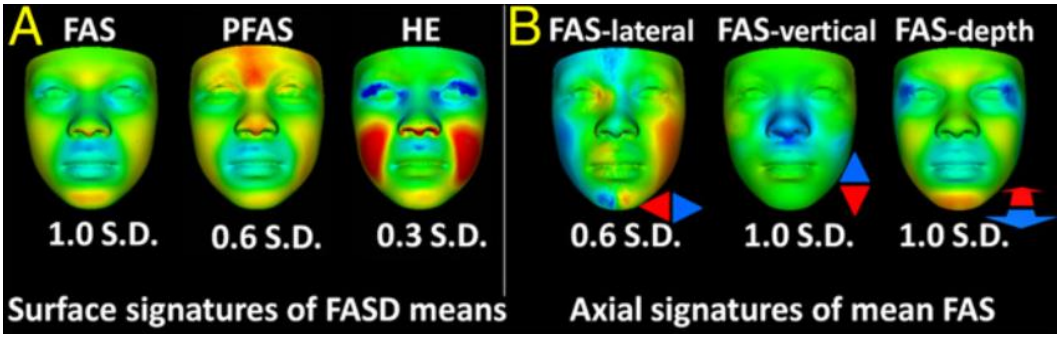


Figure 3.4 Signature heat maps for the average faces of each of the FASD clinical categories. A) To overcome facial growth differences, FAS, PFAS and HE means were scaled using a nasion to gnathion distance ratio prior to normalisation giving signature representations of shape only difference. Blue colouring on philtrum represents a convexity in this region indicating philtral smoothness seen in all 3 groups albeit at different SD scales. Areas of red on the chin and midface are indicative of retrognathia and midfacial hypoplasia. Dysmorphic features in the HE and PFAS groups were subtler and required lower thresholds of significance to visualise effects. B) For greater directional sensitivity we also visualise differences along the lateral, vertical and depth axes. In the lateral heat map of the average FAS face, red-blue opposing colouring on the inner and outer canthi is representative of a reduction in PFL. In the vertical axis, localised blue colouring is indicative of an upward displacement of the nose. The depth comparison shows a red hue on the chin, between the eyes, nasal bridge, and midface representing retrognathia, flattening of the nasal bridge and midfacial hypoplasia.

3.4.3 Facial Signatures Reveal Individual Dysmorphology

Figure 3.2.ii shows a frontal and profile view of a FAS individual, selected due to the characteristic presentation of FAS facial features in addition to idiosyncratic mild proptosis (anterior bulging of the eye from the orbit). The first shows the raw point to point differences in mm when compared to the match mean (n=35) at a scale of ±5mm. Green colouring indicates regions coincident with the matched-control, red refers to areas with contraction and blue – expansion. The figure on the right shows the significance of these differences to ±2 SD. A localised blue area indicates philtrum smoothness in the centre of the philtrum. By altering the sensitivity of the scale on the heat map, we can better identify dysmorphology in localised regions. Alternatively, localised models focusing on a segmented region can be built to increase localised sensitivity of DSM synthesis. Figure 3.5 shows a series of signatures from an upper-lip/philtrum model. These heat maps reveal the convexity of the philtral grooves and the resulting degree of smoothness. Applying this principle to other segmented regions of the face increases the sensitivity of DSM surface synthesis, providing finer detail for assessing localised facial dysmorphology. Region-specific models are particularly useful for identifying certain features that relate to the phenotypical traits displayed in FASDs, i.e. malar flattening, shortening of the nose, thin upper lip and shortened PFL.

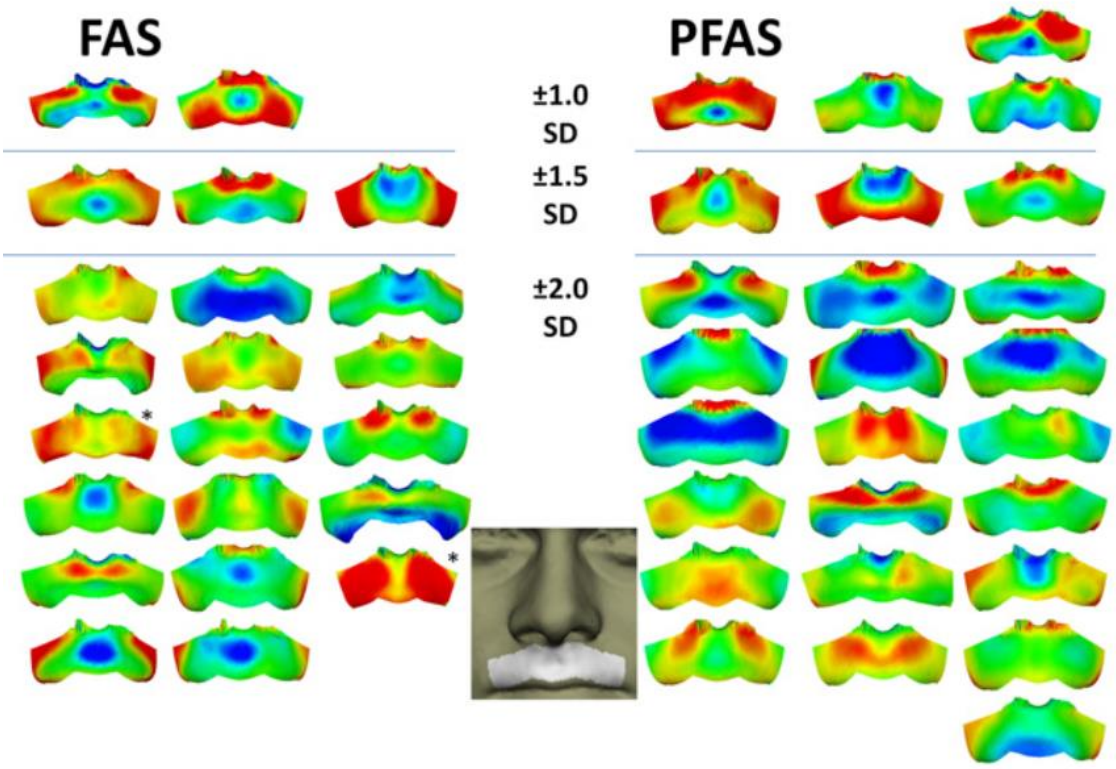


Figure 3.5. Examples of localised signatures of the upper-lip. Philtral grooves typically discriminated by regions of blue surrounded by green. A green or yellow area surrounded by red is an equivalent colouring when growth reduction is significant compared to matched controls. The centre insert shows facial region modelled.

3.4.4 Accurate Classification of FAS, PFAS and HC on Face Shape Alone

Multi-folded discrimination testing for the full face and smaller localised regions (periorbit (eyes), perioral (mouth), perinasal (nose) and mid-line profile), using 20 randomly sampled splits, determined 90% to 10% training and unseen test sets respectively for classification testing. Results for HC vs FAS and HC vs FAS+PFAS achieved high overall discrimination rates (i.e. concordance with clinical diagnosis) for the full face (FAS: 0.97–1.00), periorbit (FAS: 0.98, FAS+PFAS: 0.90), and midline facial profile (FAS: 0.93, FAS+PFAS: 0.93). For each region, we see a diminished accuracy in classification in the PFAS group, most likely due to the relaxation of criteria for growth, brain defects and neurocognitive deficits which increase the heterogeneity of effects. For each comparison, the midline facial profile achieved consistently high discrimination rates. This region encompasses a minimal surface coverage area of complex midline geometry, where foetal development is significantly affected in FASDs. An HC-FAS mean morph, visualising mean differences of the profile captures mid-facial hypoplasia, retrognathia and flattening of the philtrum (http://www.ucl.ac.uk/~sejjmfj/fasd_morphs.htm).

	HC vs FAS			**HC vs FAS+PFAS**		
	CM	**LDA**	**SVM**	**CM**	**LDA**	**SVM**
Face	0.967	0.967	1.00	0.892	0.909	0.909
Periorbit	0.983	0.917	0.967	0.892	0.900	0.892
Perioral	0.850	0.850	0.884	0.883	0.883	0.883
Perinasal	0.833	0.850	0.934	0.825	0.825	0.817
Profile	0.933	0.933	0.917	0.925	0.933	0.917

Table 3.4 Discrimination rates estimated as the mean area under the ROC curves of 20 cross-validation tests, and correspond to the probability of accurately classifying pairs of individuals; 1 taken from the HC set and the other from the exposed set. The full face in addition to sub-regions of the face were tested for HC vs FAS and HC vs FAS+PFAS comparisons.

3.4.5 Facial Signature Graphs Provide a Panorama of Dysmorphism across the Foetal Alcohol Spectrum

For each subject in the dataset, we generated facial signatures relative to their age-matched control means (n=35). Individuals with an age difference >11 months from their matched mean were discarded, omitting 16 HEs from the signature analysis. Figure 3.6 shows the remaining 107 eligible alcohol-exposed subject, in their respective FASD categories. Visually we identify a vast number of individuals with FAS, and PFAS expressing a dominant red hue on the heap maps, indicating a reduction in facial growth and localised differences. Inspecting the HE group, we see less homogeneity and convergence to green-blue colouring.

Figure 3.6 Facial signatures of the 107 alcohol-exposed individuals, at ±2D normalised against age-matched means (n=35) and divided into their respective FASD classifications; FAS, PFAS and HE. Highlighted are the 4 individuals with a PFAS diagnosis with graph nodes that repel from the FAS & PFAS clusters (cluster 18; Figure 3.8) which are shown in greater detail in Figure 3.7.

The corresponding facial signatures graph of these individuals (Figure 3.8A and Figure 3.8B) offers an indication of interconnectivity across the FASD spectrum. These graphs are colour coded for the respective FASD groups; red (FAS), blue (PFAS) and green (HE). Typically peripheral nodes will represent the greatest severity of facial dysmorphism, while central nodes will characterise those with milder features. FAS and PFAS labelled nodes predominantly cluster together, except four PFAS in cluster 18 whose facial signatures are not typical for either group. On closer inspection of the four PFAS outliers (Figure 3.7), we see more localised areas of red and generally more green and blue colouring compared to the individuals with FAS and PFAS. These signatures do not appear to display some of the typical FAS/PFAS features as per the criteria, and show little resemblance to the FAS or PFAS means.

In contrast, the HEs separated almost equally into two regions; one in a predominantly HE cluster and the other intermixed with FAS+PFAS, emphasised in Figure 3.8C by the use of green squares. With the introduction of the HC category, this relative clustering of the HE individuals

remained unchanged. Thus, a subgroup of HE individuals shows facial dysmorphism that is more FAS+PFAS-like than control-like.

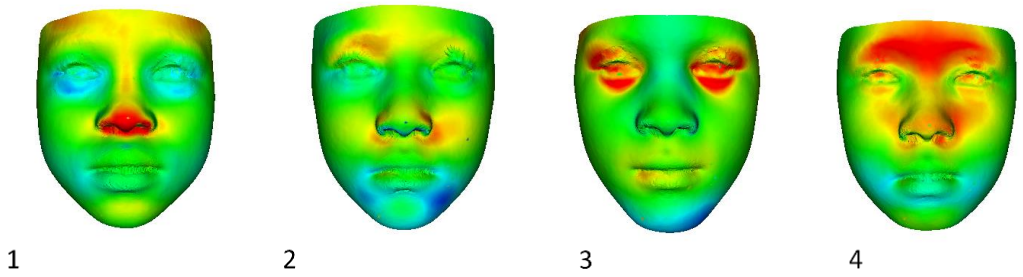


Figure 3.7 4 individuals with PFAS that repel from FAS & PFAS clusters in the exposed signature graph (cluster 18; Figure 3.8). 1, 2 and 4 show occipitofrontal reduction represented by red colouring on the forehead region and 3 shows predominantly green colouring indicating coincidence with its age-matched control.

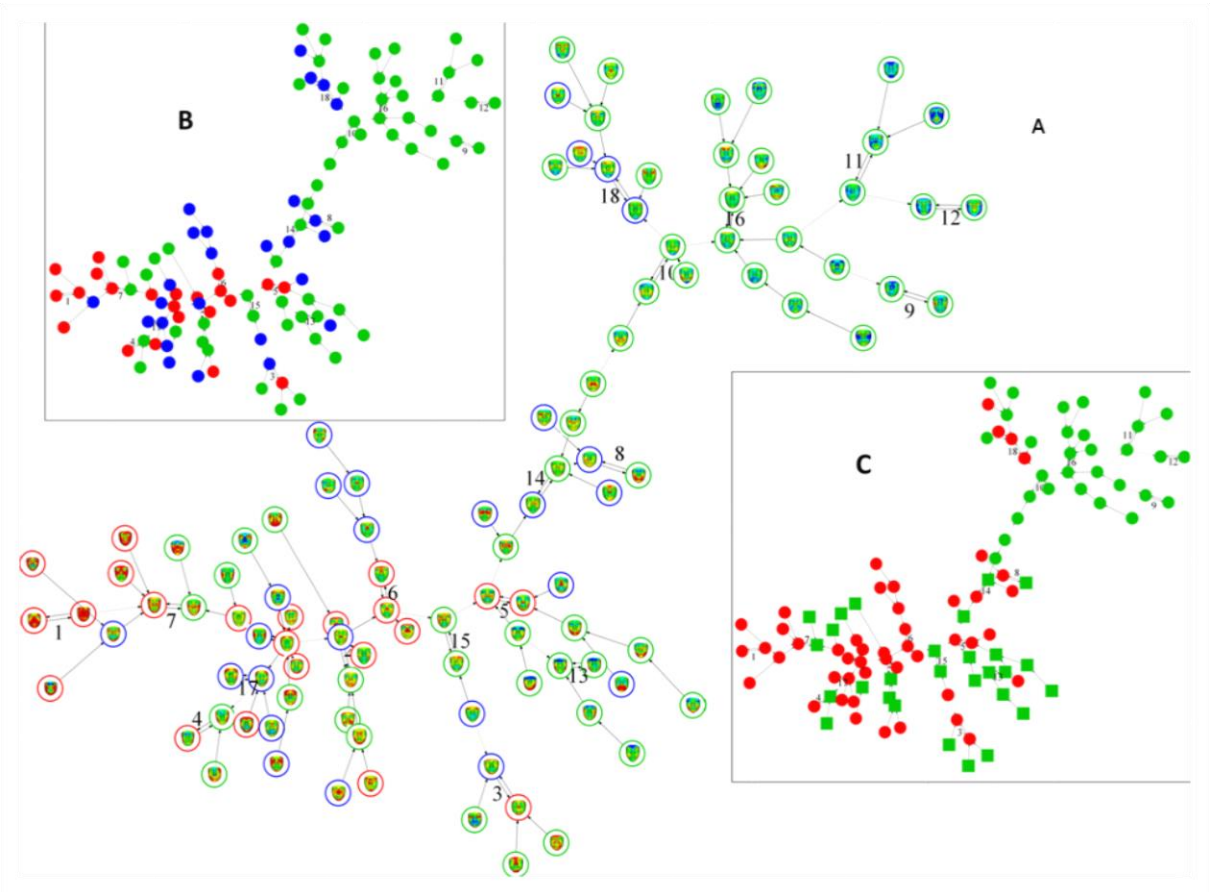


Figure 3.8 Face signature graphs for 107 alcohol-exposed children colour coded as FAS (red, n=22), PFAS (blue, n=26) and HE (green, n=59) (A) Lower part of the graph shows FAS+PFAS groups clustered with HE. The upper section shows HE clustering separately. (B) The colour coded version distinguishes FAS and PFAS C) Colour coded FAS+PFAS combined (red circles, n=48), PFAS/FAS-like HEs (green squared, n=28) and the remaining HEs (green circles, n=31) to illustrate the separation of the upper and lower clusters.

3.4.6 Subset of the HE Group Has Facial Dysmorphism That Is More FAS-like Than HC-like

Figure 3.8C shows an alternative colour coding, with HE nodes in green and FAS+PFAS in red, dividing HEs into square and circular node symbols. Squares are representative of the 28 individuals that lie below the 8th cluster who have a greater affinity to the FAS+PFAS set. The addition of the HC group forms a new signature graph FIGURE, in which the same HE subset retain their relative position amongst the FAS+PFAS nodes. Closer examination of these HE individuals revealed more FAS-PFAS like features and growth deficits. We scrutinise six individuals from this group (Figure 3.9) with mid-facial hypoplasia and/or a flat nasal bridge. These features are clear in signature graphs represented as red-yellow regions and visually in the profile views (row 2). Additionally, for a more intuitive visualisation, we generated animations showing a dynamic morph to an individual from their age-matched control mean (http://www.ucl.ac.uk/~sejjmfj/FASd_morphs.htm). Subjects 1-4 reveal mid-facial hypoplasia and philtral smoothing, 1, 2 and ± show retrognathia and in 4–6 there is an indication of upper lip thinness which is not apparent in individuals 1-3.

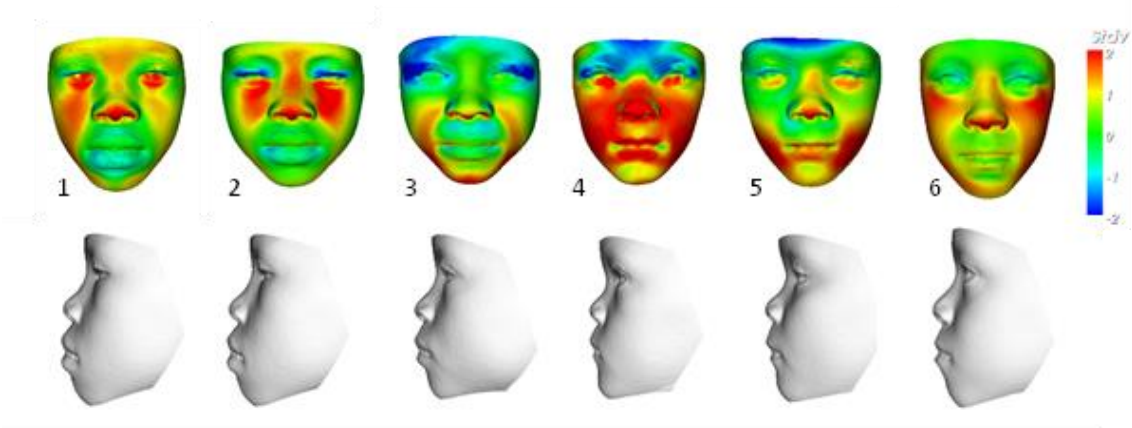


Figure 3.9 Facial dysmorphism of 6 individuals with an affinity to FAS-PFAS nodes in signature graphs of the exposed sets. Row 1 shows facial signatures at ±2SD for each individual. Dynamic morphs have been generated to better emphasise facial morphology relative to their age-matched controls (http://www.ucl.ac.uk/~sejjmfj/FASd_morphs.htm). Subjects 1 through to 4 reveal mid-facial hypoplasia and philtral smoothing, 1, 2 and 6 show retrognathia and in 4–6 there is an indication of upper lip thinness which is not apparent in individuals 1-3.

To determine levels homogeneity within the group, we generated an additional signature graph containing only HEs normalised against themselves (Figure 3.10). Visual inspection shows there to be a divided spectrum of colouring in the heat maps from one end to the other (Figure 3.10A). On the left-hand side of the graph, red-green colouring is dominant and by traversing through the nodes to the right-hand side the colouring transitions to green-blue. A binary version of this graph (Figure 3.10B), reveals the node division from the FAS-PFAS like hyper-cluster of HEs remaining perfectly within their grouping. FAS-PFAS like HE nodes are drawn with green squares, with the remaining as green circles.

3.4.7 Neurobehavioral Differences in the Cape Coloured HE Group Indicate Presence or Absence of FAS-like Facial Features

The facial signature graph of the 107 exposed subjects in Figure 3.8 identified a subset of 28 HE individuals whose facial dysmorphology had an affinity to the FAS and PFAS nodes within the graph. To reiterate, we refer to this HE partition as being more FAS-like and the remaining (with a more control-like face signature) to be control-like. By testing neurocognitive measures from we observe the FAS-like subset of HEs to be significantly diminished neurocognitive scores than the control-like group, particularly for CVLT-C and WISC-IV Verbal Comprehension IQ. Mean FAS-like HE performance distinctly resembled that of individuals with FAS, whereas the remaining control-like HEs resembled controls.

	FAS	PFAS	FAS-like HE	HC-like HE	HC	t FAS-like vs HC-like
WISC-IV[a]						
Verbal comprehension	65.4	63.0 (8.0)	65.5 (12.9)	73.3 (10.4)	73.3 (12.3)	-1.80*
CVLT-C[b]						
List a total correct	42.7	41.5 (11.4)	40.0 (11.1)	47.3 (9.0)	45.8 (9.5)	-2.02**
Recognition discrimination	88.5	88.3 (12.3)	84.3 (20.1)	93.7 (6.0)	93.2 (9.4)	-1.89*

Table 3.5 a. WISC-IV IQ Verbal Comprehension Index b. CVLT-C List A total correct. Values are mean (SD). For WISC-IV values represent standard scores and for CVLT-C the mean correct response count.
n = 9 FAS, 19 PFAS, 13 HE1, 16 HE2, 22 HC.
n = 13 FAS, 19 PFAS, 13 HE1, 18 HE2, 38 HC.
* P <0.08.
** P<.05.

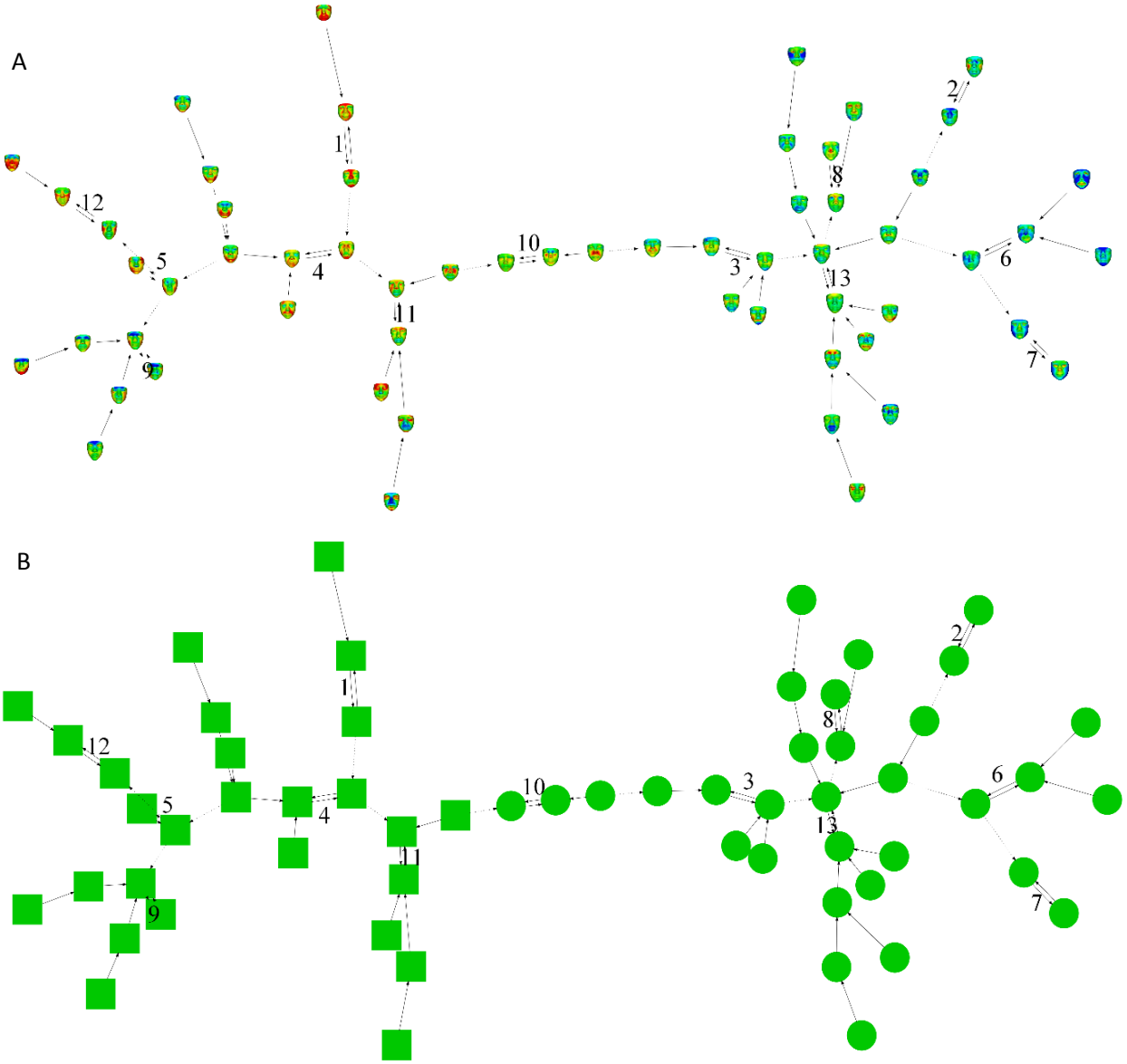


Figure 3.10 HEs (n=59) normalised against themselves. A) Green-red colouring dominates the left side nodes transitioning to green-blue on the right when visualising the heat maps of HE individuals. B) Nodes illustrated by green squares are the same HE individuals that clustered together in the FAS+PFAS+HE signatures. The grouping remains strictly one sided with the remaining HE nodes as green circles on the opposite end of the graph.

3.5 Discussion

Facial gestalt plays a key role in the identification of a FAS diagnosis. The clinical protocol for the determination of the standard phenotype is necessarily subjective, with the skill and experience of the clinician having a significance of influence on the diagnostic outcome. Milder phenotypes further down the FASD spectrum express fewer and less obvious features that can overlap with genetic syndromes. Analyses utilising computer-based recognition of facial features show promising results for the identification of FAS, but do not facilitate the full FASD full spectrum. Classification schemes in other studies assess agreement only with unstandardized clinical categorisation. Therefore, comparison of such studies should consider not only classification accuracy but also pattern-matching methods, clinical categorisation and individual facial features.

In a study of a different South African population (FAS n=17, controls n=17) of mixed ancestral background, clinical criteria included; facial dysmorphism, alcohol exposure, growth (based on height and weight), verbal or non-verbal intelligence quotient tests and a behavioural assessment (Mutsvangwa *et al.*, 2010). Images were captured using a stereo-photogrammetry capture system to extract anatomical points for landmark-based discriminant function analysis. This study evaluated shape differences at two different time points (5yrs and 12yrs) utilising drop-one-out testing to discriminate between FAS and control groups. Results indicated sensitivity-specificity scores of 100%:91% at the 5yr time point and 76%:83% at 12yrs. These results support the notion that FAS facial features are more pronounced in younger generations and subsequently diminish with age (Astley and Clarren, 1996). While landmark based study like this can accurately define geometric morphometry, its shortcoming when applied to FAS is the inability to determine surface curvature in soft tissue areas such as the philtrum.

Using laser scanning for acquisition, (Fang *et al.*, 2008) developed an approach to automatically detect FAS facial features in both a subset of our data (36 FAS and 31 controls) and a Finnish cohort (50 FAS and 32 controls). Using the same clinical criteria as our study, they employed machine learning and pattern recognition algorithms to determine a set of features that best discriminated FAS from control groups. Neural network classifiers on a 66%:33% training-unseen test set pair were applied to both ethnic populations independently and combined. This method correctly classified the Cape Coloured subset at 91.7%:90% and the Finnish cohort at 88.2%:100% sensitivity to specificity. The combined group achieved 82.75%:76.2% sensitivity to specificity. The selection of different features for each cohort suggests a variance of effect dependent on ethnicity.

An evaluation of the 4-Digit Code (Astley and Clarren, 1996) tested the ability to identify the FAS facial phenotype from 2D photos. This study involved 84 controls and 42 FAS subjects from different ethnic backgrounds, where they based clinical identification on the recognition of the facial gestalt from 2D photographs. They acquired the three key facial features; PFL, lip thickness and philtrum smoothness using a series of manually guided procedures from a series 2D photographic images. For classification testing, the dataset split into a 50:50 training-unseen test set pair, classifying using stepwise discriminant function analysis to identify facial features that best differentiated FAS from control subjects. Rather than calculating the area under the ROC curve, agreement was calculated at one value of the induced discriminant function. A combination of three objective measures were selected; the ratio of PFL to inner canthi distance, lip thinness (upper lip perimeter^2/area) referred to as circularity, and philtrum luminosity. In

addition to the subjective lip-philtrum measures derived from a 5-point Likert scale. With PFL/inner canthi ratio, philtrum smoothness and upper lip thinness measured on Likert scales, classification testing resulted in sensitivity-specify values of 100%:100%. Similarly, sensitivity-specify values when replacing the lip thinness Likert score with the circularity measure resulted in 100%:100%, and when replacing both lip and philtrum Likert scores with circularly and philtrum luminosity respectively achieved 100%:93%.

The primary aim of our study was to enhance the ability to recognise facial dysmorphology across the FASD spectrum. Ultimately the goal is to assist clinicians to accurately and objectively identify individuals at high risk for deficits resulting from PAE, from different ethnic backgrounds, and across all age ranges. We utilised multiple random unseen-test set pairs for pattern-matching and ROC-based analysis to estimate facial classification accuracy across a series of FASD diagnostic categories (Table 3.2). This DSM based representation of 3D facial shape was capable of achieving perfect agreement for FAS and near perfect agreement for FAS + PFAS combined. Notably, the midline profile - an area representing only a fraction of the available 3D surface, performed consistently well for both FAS and FAS + PFAS sets. Despite the small coverage, the profile region encompasses a significant amount of FASD associated morphology including; philtrum smoothness, philtrum length, flat nasal bridge, prognathism and anteverted nares, as documented in the updated IOM guidelines for diagnosis (Hoyme *et al.*, 2016). Midline facial effects could also relate to the disproportionate effect that alcohol as a teratogen has on midline brain anatomy (Riley *et al.*, 1995b; Bookstein *et al.*, 2007; Yang, Phillips, *et al.*, 2012; Jacobson *et al.*, 2017). Given this observation, further work within this thesis will investigate midline defects across the FASD spectrum utilising this profile region.

In this chapter, we showed how FASD-control comparisons could be visualised using heat map comparisons and dynamic morphs to reveal facial dysmorphism that may otherwise be overlooked. Additionally, we built signature graphs linking individuals based only on facial dysmorphology, ignoring clinical classification. A signature graph of 107 exposed subjects (FAS, PFAS and HE) provided a panorama of dysmorphism across the FASD spectrum. FAS and PFAS nodes within the graphs clustered almost entirely with 28 individuals of the HE set (except 4 individuals with a PFAS label outlined in Figure 3.7). Furthermore, when testing the neurocognitive variables, the FAS-like HE subset performed more like FAS compared to the remaining HEs, who's measures were closer to that of controls. When HE individuals were isolated in a graph and normalised against themselves, the same 28 nodes clustered together. While this subset showed FAS-PFAS like affinity in the graphs, individuals showed few of the classic FAS features. Instead, features such as malar flattening, flat nasal bridge, micrognathia and retrognathia were recognisable from heat maps and dynamic morphs between control and individuals. These features are not used in criteria for assigning FASD diagnoses but correspond to a recently revised list of minor anomalies outlined in the IOM guidelines (Hoyme *et al.*, 2016). Following these guidelines, dysmorphic variance in an individual is calculated using an objective scoring system with measures of growth and minor anomalies. Features conforming to our observations of the HE FAS-like subset include; hypoplastic midface, flat nasal bridge, prognathism and anteverted nares.

3.6 Conclusions

In this chapter we have identified a novel approach for the identification of facial effects associated with PAE. Normalised heat maps and dynamic morphs of FASD individuals deliver an

enhanced form of feature recognition and an objective method of facial assessment. Additionally, the ability to classify FAS and PFAS subjects based on full and isolated regions of the face provided encouraging results. These clinical categories are well suited to this form of discrimination testing given the homogeneity of facial features used for clinical assessment. Given the lack of typical features, classification methods will not apply in the same manner for HEs. Instead, facial signature graphs were able to select distinctive HE subsets, showing some heterogeneity of facial and neurocognitive effects in this vastly heterogeneous group. It is possible some features may go unrecorded during routine physical screening, either due to their subtlety or because of the underestimation of secondary features. This novel methodology for assessing and identifying alcohol-exposed facial phenotypes shows promise for an objective diagnosis across the foetal alcohol spectrum.

4 Ethnic Differences In the Effects of Prenatal Alcohol Exposure

In this chapter, we introduce novel techniques of facial assessment using facial curvature to recognise the effects of prenatal alcohol exposure across different ethnic groups.

4.1 Introduction

In the previous chapter, we assessed a South African Cape Coloured population, applying DSM analyses to examine individual dysmorphology and discriminate healthy controls from severely affected individuals with FAS and PFAS. We also utilised signature graphs to identify FAS-like similarity within the heavily exposed population who do not meet clinical criteria. In this chapter, we compare the same Cape Coloured cohort and introduce a Caucasian cohort recruited within the CIFASD consortium. Comparing control-FAS discrimination results using region-based models uncovered inconsistencies between the two ethnic groups. We also introduce facial curvature as an innovative application in FASD facial analysis to demonstrate its ability for feature recognition, and to explain some of the observed ethnic inconsistencies. Additionally, we show how some individuals who do not fall within guidelines for diagnosis can give rise to orbital hypertelorism. This result appears concordant with the long-standing notion that alcohol exposure at a particular time point during gestation causes an increase in brain hemisphere separation, inevitably causing orbital separation (Sulik et al., 1986). Results within this chapter originate from a recently published study (Suttie *et al.*, 2017), in which data collection for the Caucasian cohort was undertaken by members of the CIFASD consortium and analysed under the supervision of the principal investigator Prof Peter Hammond.

Diagnostic categories used in this chapter differ only minimally from those from the Cape Coloured cohort with the exclusion of the partial foetal alcohol syndrome (PFAS) label. Leaving three categories; foetal alcohol syndrome (FAS), non-syndromal heavily exposed (HE) and non-

exposed healthy controls (HC). This analysis focuses on utilising local sections of the face given the accuracy of control-FAS discrimination found in the Cape Coloured population using this approach (Chapter 3). In particular, we previously identified that the mid-facial profile region; a region of minimal coverage expressing the profile of the face, proved to contain a quantity of information suitable for accurate FAS and PFAS discrimination.

4.2 Study Populations

Excluding PFAS from the previously mentioned Cape Coloured cohort studied in Chapter.3 left 166 individuals (age range; 6.8 to 16.2 yrs). The CIFASD consortium recruited 249 Caucasian participants across four sites in the United States (n=207 from; Minnesota, San Diago, Atlanta and Los Angeles) and from several countries in Europe (n=42). We excluded participants who self-reported as African-American, Native-American, Asian and Hispanic to maintain uniformity. Due to the scarcity of FAS diagnosed individuals in the data available we could not analyse these additional racial groups. Diagnostic categorisation did not use the PFAS label and Cape Coloured:Caucasian diagnostic category sizes were as follows: HC (69:141); FAS (22:35); HE (75:73).

4.3 Methods

4.3.1 Clinical Assessment

Table 3.2 from the previous chapter outlines criteria used to differentiate FAS, PFAS and non-syndromal HE individuals. For each cohort, multiple specialised dysmorphologists examined individuals, blinded to prenatal alcohol exposure history, to identify FAS facial features and associated growth defects. Clinicians utilised standard protocols for measurements and diagnostic labelling (Hoyme *et al.*, 2005) employing the Astley and Clarren Lip-Philtrum Guide (Astley and Clarren, 2000) for upper lip vermilion and philtrum assessment. Palpebral fissure length (PFL) was measured using a manually positioned ruler. We excluded any Individuals with a known genetic disorder. Examining dysmorphologists made final decisions on FASD categorisation based on an agreement by consensus. Alcohol intake of mothers during pregnancy for the children of each ethnic cohort was both prospectively and retrospectively recorded. Mothers who reported complete abstinence or very minimal alcohol intake were designated as controls unless they met criteria for FAS.

4.3.2 Face Analysis

Both cohorts used similar static, tripod-mounted commercial 3D cameras from 3DMD *(www.3dMd.com)* for the acquisition of the 3D facial images. 180-degree ear-to-ear 3D images with an order of 25,000 mesh points, captured at sub-millimetre accuracy (geometric resolution <0.2mm) provided an accurate 3D facial surface representation. Each participant had their 3D photo taken as part of the clinical assessment. Using in-house software (Facemark), the author manually annotated a sparse set of 24 reliable anatomical landmarks (Gwilliam, Cunningham and Hutton, 2006) to each 3D surface (Chapter 3; Figure 3.1). Landmarks were used to Procrustes align the facial surfaces and induce a dense correspondence of points for DSM model building. We computed DSMs for the whole face and seven localised regions of the face (Figure 4.1) using in-house software (Hammond and Suttie, 2012).

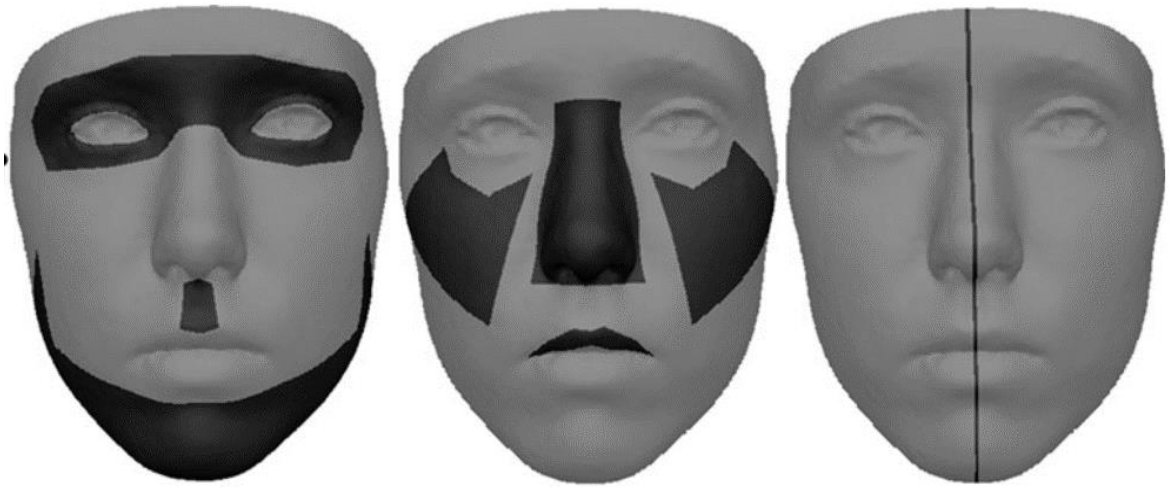


Figure 4.1 Facial regions used for DSM model building. Left; periorbit, philtrum, mandible. Centre; perinasal, malar region, upper lip vermillion. Right; midline profile.

Separately testing for both ethnic groups, we applied multi-folded classification techniques; closest mean (CM), support vector machines (SVM) and linear discriminant analysis (LDA) in a supervised learning fashion to classify individuals and groups previously categorised clinically. We discriminate between sets of unseen randomly sampled test examples across 20 random 90% to 10% training-unseen test pairs. We calculate discrimination rates between two subgroups (control-FAS subsets) in terms of the mean area under the receiver operating curves (ROC) of the 20 trials, representing the probability of correctly classifying a pair of randomly selected faces with each of the pairs from a different subgroup. Applying this approach to the full face and seven isolated regions; periorbit, philtrum, mandible, perinasal, malar region, upper lip vermilion and mid-line profile (Figure 4.1) provided localised HC-FAS discrimination rates. The Cape Coloured analysis from the previous chapter excluded the philtrum, malar region and upper lip vermilion. Furthermore, in this study, we exclude the perioral region and replace it with a localised upper lip vermilion patch.

4.3.3 Face Signatures and Signature Graphs

We similarly analyse group and individual dysmorphism as in Chapter 3, using *face signatures* and *face signature graphs* (Hammond and Suttie, 2012). Ethnicity-age matched means of 35 contiguous controls were created for each to calculate their facial signatures. Facial signatures of means of each clinical subgroup were also generated producing heat maps of the normalised surface to normal shape difference in addition to directional displacement parallel to lateral, vertical, and depth axes.

To determine linkage within each ethnic cohort, we generate facial signature graphs similar to those illustrated in Chapter 3. Signature graphs with nodes labelled by colour with clinical categories can be “collapsed” into a simplified form. Connected vertices of the same group (colour) are combined to form a super-vertex while preserving connectivity to differently categorised vertices. Nodes on the signature graph represent these collapsed vertices, sized relative to the number of vertices that form them and respectively numbered. A measure of entropy, referred to as a dispersion index, summarises the associated clustering, or dispersion, of the nodes of each labelled group (Hammond, Suttie, *et al.*, 2012). Lower dispersion indices

represent stronger similarity or homogeneity within the category, with a dispersion index of 0 meaning that all members of a group will only connect to each other indicating complete homogeneity (relative to the dataset). In contrast, a dispersion index of 1 signifies complete dispersion of a group with no inter-connected vertices within that labelled group as they lack homogeneity. If only two groups exist within a graph (e.g., FAS and control for the same cohort, or Cape Coloured and Caucasian for the same diagnosis), dispersion indices will represent relative homogeneity of each grouping and the degree of similarity regarding normalised difference from match controls. A full explanation of collapsed graphs and calculating dispersion indices can be found in Appendix 2.

4.3.4 Curvature

The curvature of the face is defined relative to the underlying surface mesh of points of the 3D surface. The algorithm employed is derived from the VTK library (www.vtk.org) and calculates mean curvature **H** at a mesh point **p,** as a weighted average over the edges between each pair of connecting cells adjacent to surface point **p.**

$$\mathrm{H(p)} = \frac{1}{\mid \mathrm{E(p)} \mid} \sum_{\mathrm{e} \in \mathrm{E(p)}} \mathrm{length(e)} * \mathrm{angle(e)}$$

where **E(p)** is the set of all adjacent edges at point pp, and **angle(e)** is the angle of edge ***e*** at point **p** The contribution of every edge is weighted by **length(e).** Curvature can be represented as a heat map, with convex/protruding features, such as the tip of the nose coloured in blue, and regions of concavity such as the philtrum or canthi of the eyes coloured in red. Rough localised surface geometry from image capture causes isolated curvature noise (facets coloured with extreme red or blue). We overcome this to some degree by applying smoothing to the curvature values. A single iteration of a simple unweighted sliding-average smoothing algorithm flattens out minor fluctuations in curvature, providing a significantly improved curvature representation. In testing, applying 2 or more iterations appeared to diminish the level of curvature precision particularly around the philtrum area, so a single iteration is sufficient, even though some isolat0ed noise persists.

We can also modify this calculation to determine curvature about an axis orthogonal to a plane in which the curvature falls. Two specific axial curvatures are particularly useful for assessing the concavity of the philtrum. Some individuals have a philtrum that has a deep vertical groove and others have little in the way of a vertical groove, but have a curl-like shape. We use the terminology “curl” and “groove” to distinguish between them and a right hand “rule” as a convenient way to define them (Figure 4.2). The facial landmarks left, and right exocanthi are used to define a horizontal axis relative to the face surface, and with the thumb in the direction of the axis, the fingers follow the curl we are trying to discriminate. We use an axis defined by the facial landmarks subnasale and upper lip centre to identify curvature that would form the normal “vertical” groove of the philtrum.

We attain axial groove and curl calculation by modifying the original mean curvature calculation to include the contribution of a cosine similarity weighting $\boldsymbol{axis}(\boldsymbol{e})$

$$\mathrm{H(p)} = \frac{1}{|\ \mathrm{E(p)}\ |} \sum_{\mathrm{e}\in\mathrm{E(p)}} \mathrm{length(e)} * \mathrm{angle(e)} * axis(e)$$

where $\boldsymbol{axis}(\boldsymbol{e})$ is calculated from the dot product of the vectors $\boldsymbol{e}$ and the user-defined $\boldsymbol{axis}$ derived from 4 anatomical landmarks. We demonstrate the efficacy of groove and curl curvature in Figure 4.3 to discriminate philtrum shape. Four examples expressing varying degrees of curl and groove curvature are shown in Figure 4.3 overlaid with heat maps of raw and normalised curvature. The first low-curl, low-groove philtrum, shows minimal colouring for both groove and curl, but when normalised, it becomes evident this shape is significantly difference from the ethnicity-age-match control norm. Examples of high-groove curvature (2nd and 4th column) shows a particularly useful application for the identification of the philtral pillars, which can be undistinguishable on smooth philtrums (low-groove curvature). Additionally, high-curl curvature can identify areas of deep concavity in localised regions or across the entire philtrum which in some exposed individuals can be "ski-jump" shaped.

Calculating mean curvature for the assessment of the philtrum is the preferred method over Gaussian or principal curvatures alone (illustrations in row 1, Figure 4.2). Principal curvatures at mesh point **p** are expressed as the minimum and maximum values of curvature at that point. Maximum curvature on the philtrum may provide an indication of the prominence of the philtrial pillars, however, will fail to define the presence of any groove. Likewise, the minimum principal curvature can delineate the groove but fails to define the philtrum pillars. Gaussian curvature is the product of the principal curvatures, and also fails in this application as its representation is dependent on the distances on the surface and not its relative position in Euclidean space. For this reason, our philtrum segmentation with Gaussian curvature does not provide any definitive curvature patterns to illustrate either the philtrum pillars or the groove.

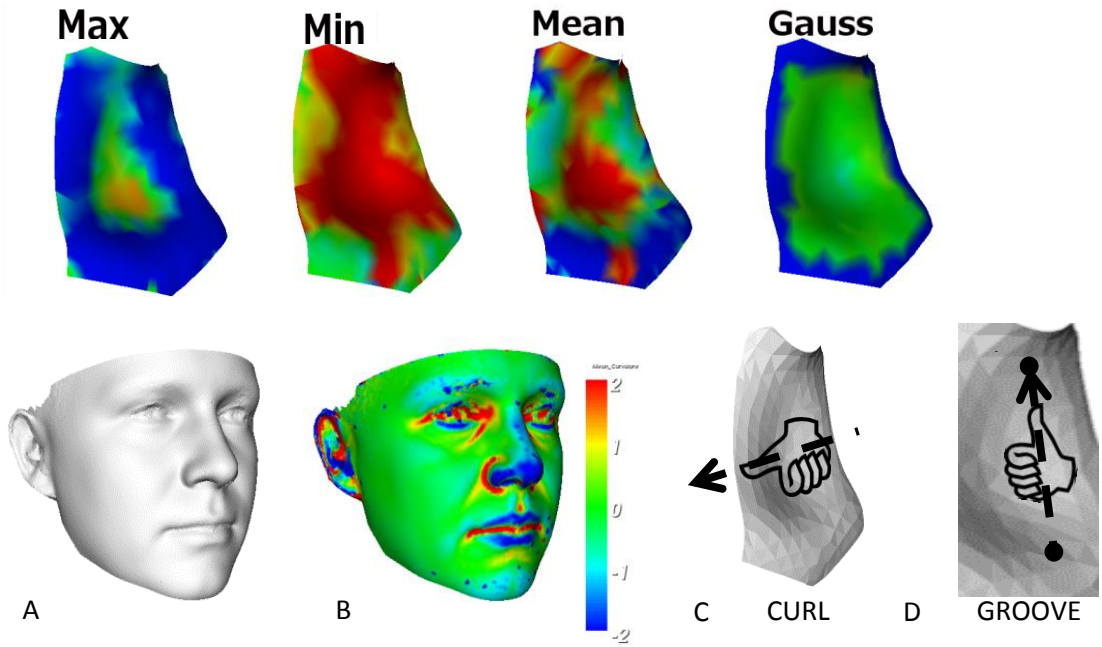


Figure 4.2. Row 1: Examples of principal (min and max), mean and Gaussian curvatures on a philtrum surface. Row 2: A & B The heat map visualizes average curvature at points across the face surface relative to the scale shown where green indicates no or little curvature at a point, red indicates inward curvature and blue indicates outward curvature. Thus, prominent features such as the nose tend to be blue and grooves or inward creases such as the philtrum are shown in red. Isolated deeply coloured surface triangles are likely to be caused by noise generated during image capture. C & D a right handed "rule"

delineates curvature of the philtrum in a plane orthogonal to the axis along which the thumb points and fingers indicate curve direction; “curl” is used to describe curvature relative to a horizontal axis (defined by the exocanthi) and “groove” is used to describe indentation relative to a “vertical” axis (defined by the line joining subnasale and upper lip centre).

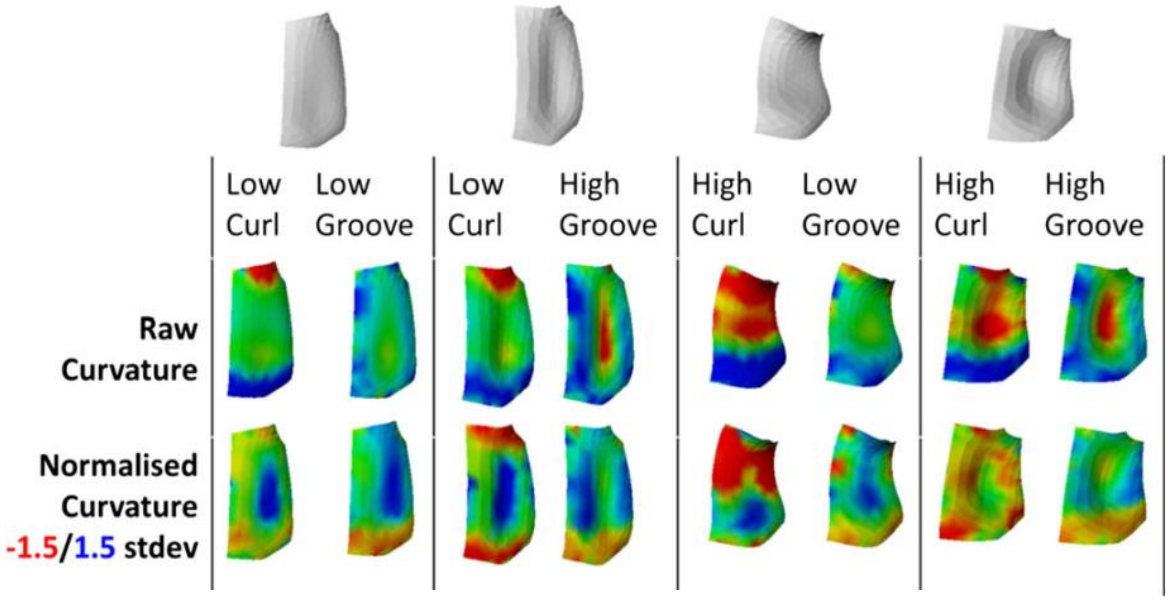


Figure 4.3 Four examples of philtrum curl and groove curvature. Row 1) The original philtrum surface, row 2) degree of curl and groove curvature, row 3) heat maps showing raw curvature, row 4) curvature normalised against 35 age-ethnicity-matched controls at ±1.5SD where; blue-red curl=convexity-concavity orthogonal to X-axis, blue-red groove=convexity-concavity orthogonal to Y-axis.

Curvature heat maps are often an improvement on displacement signatures for identifying philtrum smoothness. Facial signatures representing displacement for a FAS individual signify size as well as shape, often localised red colouring representing 'inward’ displacement, will swamp features across the face, making ‘outward’ displacement indistinguishable on a smooth philtrum.

Figure 4.4 shows the philtrum groove curvature signatures for all individuals with FAS in the Cape Coloured and Caucasian cohorts, normalised against 35 ethnicity-age-matched controls, and the corresponding collapsed signature graph they produce. Here we see a greater level of homogeneity of the Caucasian philtrum groove signatures compared to the Cape Coloured. The collapsed graph shows the Caucasian set to cluster into one large collapsed node (containing 32 individuals) with only two disconnected nodes on the periphery, whereas the Cape clusters collapse to 2 divided nodes with seven disconnected singletons. Cape Coloured and Caucasian dispersion indices reflect the level of homogeneity within each set, with values of 0.54 and 0.10 respectively. See Appendix 2 for dispersion index calculation.

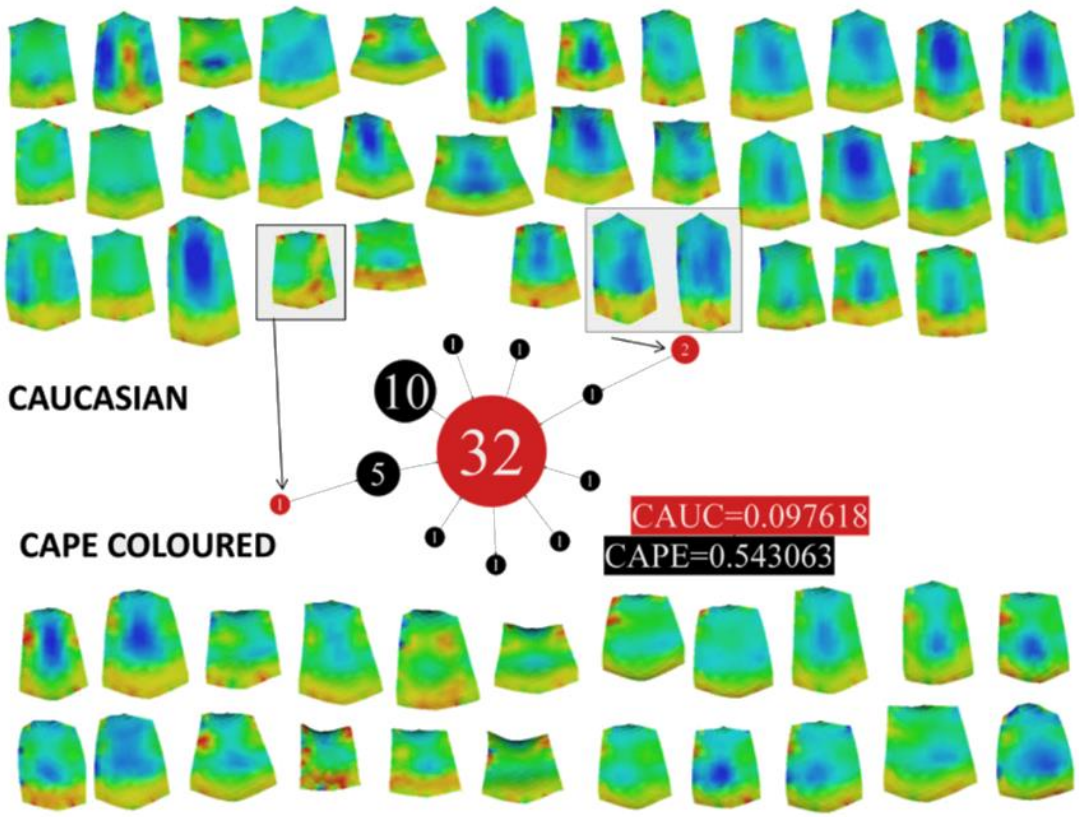


Figure 4.4 Groove curvature signatures of Cape and Caucasian FAS, and the 'collapsed' signature graph they form. Greater homogeneity of the Caucasian philtrum groove signatures is evident on visual inspection and confirmed by the collapsed graph. With the exception of 3 signatures, the Caucasian cohort cluster almost perfectly with a large cluster containing 32 of the 35 individuals with FAS, while the Cape separates into several smaller sub-clusters represented by dispersion indices: Caucasian (0.10) and Cape Coloured (0.54). See Appendix 2 for an explanation of collapsed signature graphs.

4.4 Results

4.4.1 Cohort Characteristics

Table 4.1 shows comparisons between Cape Coloured and Caucasians across FAS, HE and HC subsets for variations in gender, age, body mass index (BMI), PFL, and lip-philtrum scores. Only BMI was found to have any significant difference, with the Caucasian HEs having greater BMI than those of the Cape Coloured HEs ($p < .01$).

	SEX		AGE		BMI		PFL		VERMILION		PHILTRUM	
	F,%	M,%	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
HC												
CAPE	50.7	49.3	10.1	2.6	18.4	4.1	2.6	0.4	3.1	0.6	2.9	0.6
CAUC	41.1	58.9	11.5	3.2	19.2	3.9	2.6	0.3	3.0	0.5	3.0	0.4
FAS												
CAPE	45.5	54.5	10.6	2.9	15.7	1.6	2.4	0.2	3.8	0.7	4.0	0.7
CAUC	60.0	40.0	11.8	3.4	16.4	2.4	2.3	0.2	3.9	0.6	3.9	0.6
HE												
CAPE	52.0	48.0	10.4	2.7	17.1*	2.7	2.6	0.2	3.1	0.8	3.0	0.8
CAUC	39.7	60.3	12.1	3.0	18.9*	4.4	2.5	0.2	3.1	0.5	4.0	0.5

Table 4.1. Sample characteristics for both ethnic groups. Comparing sex, age, body mass index (BMI), palpebral fissure length, lip-vermilion and lip-philtrum scores. BMI of HEs is the only significant difference ($p < 0.01$) observed.

4.4.2 Facial Growth

We concluded in Chapter 3 that the reduction in facial growth present for the Cape Coloured cohort is significantly reduced compared to controls. Values presented below (Figure 4.5) vary slightly for Cape Coloured results as we built separate cohort-specific DSMs of the face, omitting individuals with PFAS meaning PC values differ. Similar to previous results, the facial growth of HE and HC individuals were close in both cohorts. For FAS, there was a significant reduction in facial growth compared to HC in both cohorts, but no significant difference between ethnicities. For each cohort, we constructed DSMs containing FAS, HE and HCs for which principal component 1 (PC1), in both DSMs, represented overall facial growth (Figure 4.5). Raw PC1 values for HEs show them to be similar to controls but with consistent marginal reduction. To compare FAS-HC growth between cohorts, we removed HE individuals and re-built DSMs to ensure there was no influence on the comparison. PC1 values normalised for age and ethnicity, in each of the FAS + HC only models, are still representative of facial growth. Box plot comparisons (Figure 4.6) show age-ethnicity corrected PC1 values confirming that although facial growth in subjects with a FAS diagnosis significantly reduced in both cohorts; Cape Coloured ($p < 10^{-12}$) and Caucasian ($p < 10^{-16}$), there were no significant differences for the FAS subsets between cohorts ($p = 0.32$).

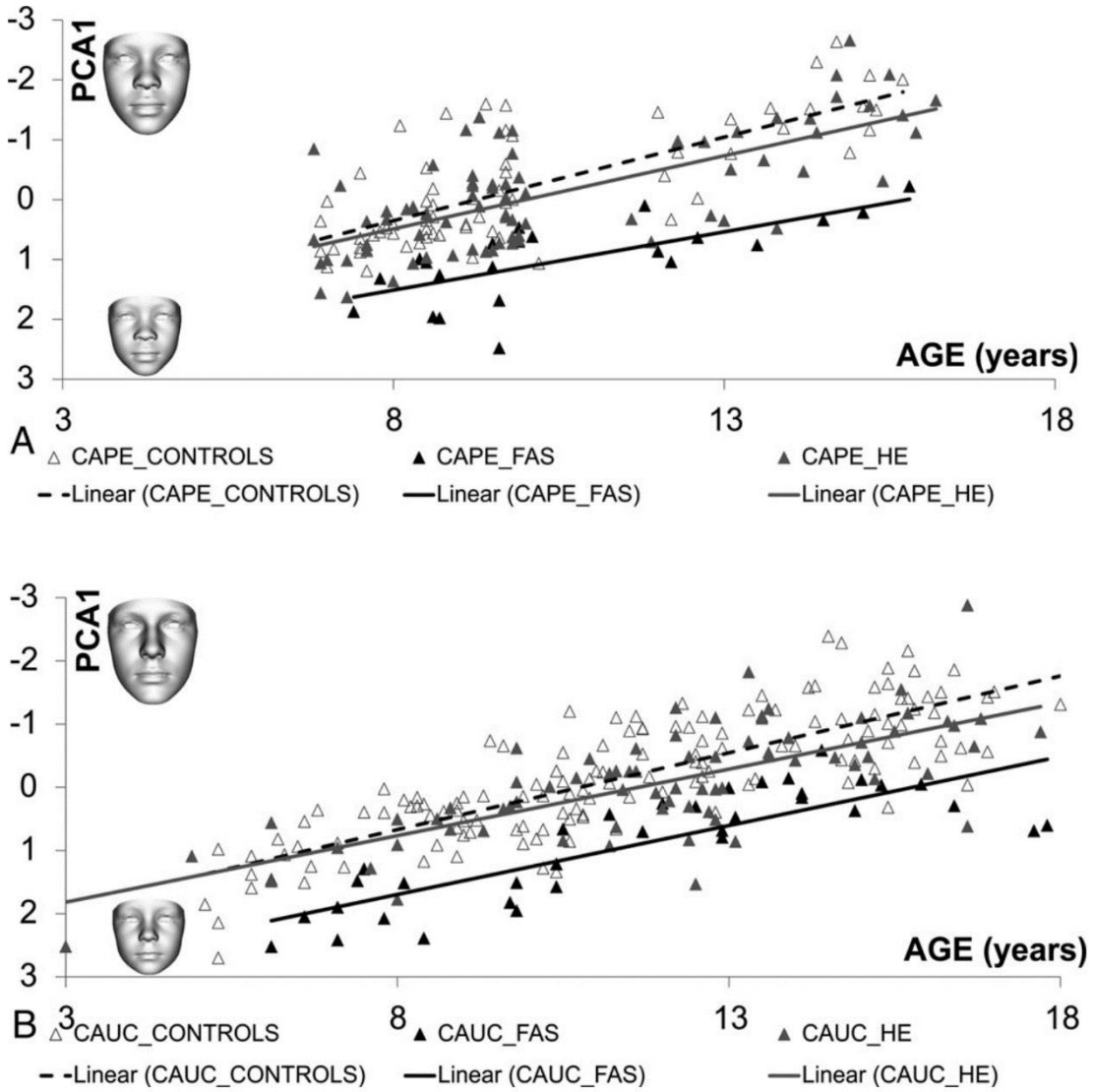


Figure 4.5 Principal component 1 vs AGE for control, FAS and HE individuals in A) Cape Coloured and B) Caucasian cohorts.

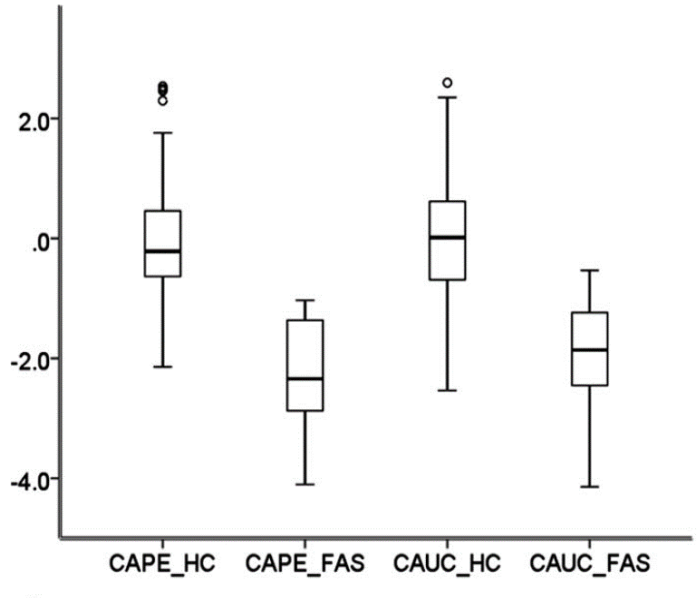


Figure 4.6. Box plot showing age-ethnicity-normalised PCA1 values representing facial growth for Cape Coloured and Caucasian FAS and controls. PC1 values are significantly reduced in both cohorts between FAS and controls (Caucasian; ($p < 10^{-16}$), Cape Coloured ($p < 10^{-12}$)) however, no significant difference exsits in FAS between the two ($p = 0.32$).

4.4.3 Control vs FAS Classification Testing

In the previous chapter, we tested agreement of control-FAS discrimination based only on surface shape using CM, SVM and LDA algorithms. Both cohorts were similarly tested using a series of midline regions (profile, philtrum, nose and upper-lip vermilions) and non-midline regions (face, eyes, mandible and malar). Table 4.2 contains discrimination rates, for all regions, estimated as the mean area under the ROC curves of 20 cross-validation tests for Cape Coloured and Caucasian cohorts. The consistency of classification rates between cohorts was higher in non-midline regions of the face, whereas midline regions were less consistent. In the Cape Coloured population, the malar performed better (0.94) than the classic FASD associated facial regions of the philtrum (0.70) and upper lip vermilion (0.73), and as previously stated the mid-line profile is particularly discriminative (0.95). In contrast, the midline profile of the Caucasian cohort was one of the weakest regions (0.83), but the face (0.95), eyes (0.92), nose (0.94) and malar (0.90) performed well. Figure 4.7 shows a simplified arrangement of these results and the relationship/consistency between the two cohorts. In the following sections we attempt to address the cohort-based discrepancies that exist in these results.

	CM		LDA		SVM	
	Cape	Cauc	Cape	Cauc	Cape	Cauc
Face	0.96	0.95	0.96	0.96	0.98	0.96
Profile	0.95	0.83	0.96	0.82	0.94	0.86
Eyes	0.95	0.92	0.95	0.93	0.94	0.92
Malar	0.94	0.90	0.95	0.91	0.95	0.91
Mandible	0.88	0.86	0.90	0.88	0.93	0.85
Nose	0.86	0.94	0.87	0.94	0.89	0.95
Lip vermilion	0.73	0.70	0.80	0.69	0.84	0.73
Philtrum	0.70	0.83	0.70	0.85	0.76	0.90

Table 4.2 Agreement of face-based classification and clinical diagnosis for Control-FAS agreement. Discrimination rates estimated as the mean area under the ROC curves of 20 cross-validation tests, and correspond to the probability of accurately classifying pairs of individuals; 1 taken from the control set and the other from the FAS set.

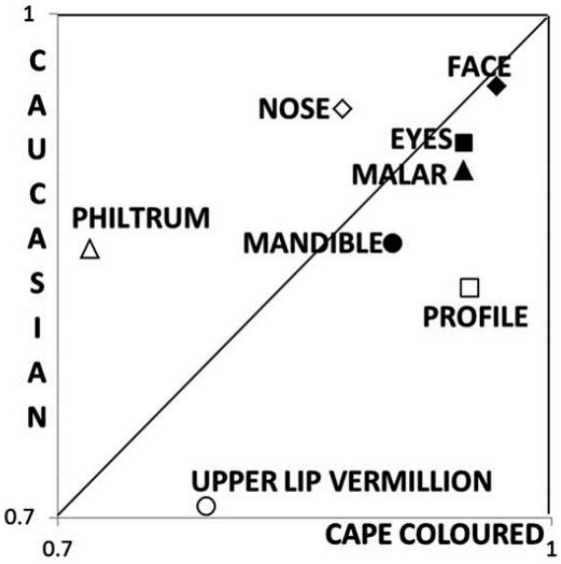


Figure 4.7. Comparison of mean classification results between each cohort for the face and 7 sub-regions. Classification consistency between cohorts is higher in non-midline regions (face, eyes, malar and mandible) compared to the midline regions (philtrum, nose, profile and upper lip vermillion).

4.4.4 Caucasian-Cape Coloured Control-FAS Classification Inconsistencies

Philtrum Differences Between Cape Coloured and Caucasian

Comparing the mean shape of the HC philtrum between the cohorts (Figure 4.8 A, F), we observe the upper half of the Cape Coloured philtrum to be flatter than the Caucasians. Generation dynamic morphs between the two means provide a visual confirmation of this difference. Normalising the Cape Coloured control mean against age-matched Caucasian controls, shows greater concavity localised to the area of the subnasale (superior region in Figure 4.8B) and greater philtrum groove flatness (central regions of blue for both curl and groove in Figure 4.8 B, C). From this, we conclude that on average the superior region of the philtrum is smoother in Cape Coloured Control than in Caucasians.

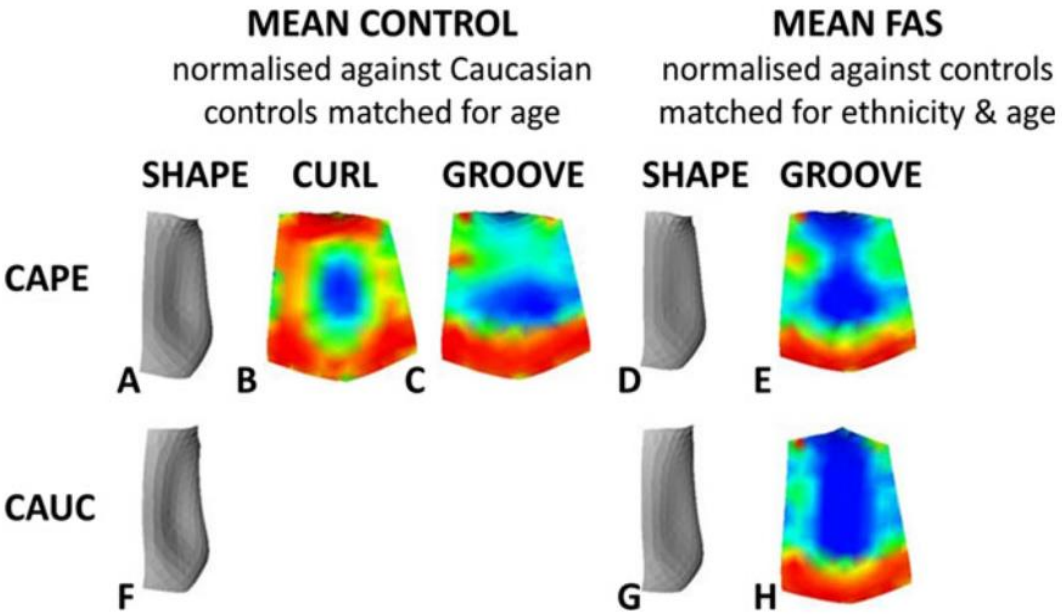


Figure 4.8 A,D,F,G) Mean FAS and control philtrum shapes of Cape and Caucasian B,C) Curl and groove curvature heat maps for Cape Coloured normalised against age-matched Caucasian controls. E,H) Both cohorts normalised against their respective age-ethnicity-matched controls. Deepest red ≤ 1.0SD; deepest blue ≥ 1.0SD.

Figure 4.8 E and H show the mean FAS philtrums for each cohort, normalised against age-ethnicity-matched controls. Regions of extreme convexity (blue) for groove curvature on the Caucasian FAS average highlight a greater degree of flattening than that of the Cape Coloured. Thus, the effect of PAE on philtral smoothing appears to be more substantial in the Caucasian cohort. By building DSMs containing both cohorts combined, we can construct signature graphs for all FAS or control individuals to test shape homogeneity. We generated facial signatures of face and philtrum, normalised against age-ethnicity-matched controls, and signature graphs collapsed to provide dispersion indices for each respective cohort for both displacement and curvature (an example of which is shown in the introduction section of this chapter; Figure 4.4). Table 4.3 shows FAS and control dispersion indices for the face and philtrum comparing each cohort. Refer to Appendix 3 to view associated signature graphs.

	FACE				PHILTRUM			
	Displacement		**Curvature**		**Displacement**		**Groove Curvature**	
	Cauc	**Cape**	**Cauc**	**Cape**	**Cauc**	**Cape**	**Cauc**	**Cape**
Control	0.80±0.03	0.69±0.06	0.58±0.02	0.77±0.02	0.69±0.02	0.79±0.01	0.19±0.02	0.59±0.02
FAS	0.76±0.02	0.50±0.03	0.70±0.01	0.28±0.03	0.56±0.02	0.75±0.02	0.07±0.01	0.55±0.04

Table 4.3 FAS and control, face and philtrum, displacement and curvature average dispersion indices and 95%CI for Cape Coloured and Caucasian cohorts. The first pair of indices in bold (0.50 and 0.28) in the FAS row and their respective control values (0.69 and 0.77) suggest that PAE reduces signature graph dispersion indicating greater homogeneity for the Cape Coloured face. Furthermore, the second bold pair (0.56 and 0.07) in the FAS row and their respective control values (0.69 and 0.19) suggest that PAE contributes towards a more homogenous philtrum shape and groove curvature in Caucasians.

The first two dispersion indices expressed in bold in the FAS row of Table 4.3 (0.50 and 0.28) and their respective control values (0.69 and 0.77), show how PAE reduces dispersion of the displacement and curvature signature graph, translating to an increase in overall homogeneity of the face in the Cape Coloured cohort. For the philtrum graphs, Caucasian FAS dispersion indices for displacement (0.56) and groove curvature (0.07) were reduced compared to their control counterparts (0.69 and 0.19 respectively). In contrast, the corresponding Cape Coloured FAS dispersion indices (0.75 and 0.55) barely differed from their control counterparts (0.79 and 0.59). This result indicates greater homogeneity of the Caucasian FAS philtrums, which we see in the groove signature graph (Figure 4.4) where 32 of the 35 Caucasian individuals with FAS form one large sub-cluster subsequently accompanied by a low dispersion index (0.10). The philtrum results above are likely to contribute towards the improved agreement for Caucasian control-FAS classification.

Midline Profile Differences Between Cape Coloured and Caucasian

A curl signature graph combining cohorts for the midline profile revealed dispersion indices of 0.28±0.01 and 0.73±0.01 for Caucasian and Cape Coloured controls respectively. The disparity of dispersion indices signifies that normalised curvature of the midline profile is more homogenous in Caucasians than in Cape Coloureds. The corresponding FAS dispersion indices revealed an opposite effect, with values of 0.48±0.03 for Caucasians and 0.31±0.02 for Cape Coloureds. These results suggest there is greater uniformity of the difference from controls in the midline profile of Cape Coloureds than in Caucasians. We measure the degree of curvature induced by PAE by calculating the log transformation of curl signature weight. By doing so, we observe a significantly greater curvature in the Cape Coloured cohort compared to the Caucasians (Cape mean: 1.49; Caucasian mean: 1.42; $p < 0.05$). The midline profile results are likely to contribute towards an improved agreement for Cape Coloured control-FAS classification compared to the Caucasian equivalent.

4.4.5 Face Signature Graph Differences between Cape Coloured and Caucasian

In the previous chapter, the signature graph containing exposed individuals (FAS, PFAS & HEs) partitioned HEs into FAS-like and control-like subgroups based on dysmorphology. Because the Caucasian cohort does not distinguish individuals PFAS, we retested the Cape Coloured signature graph omitting PFAS. Encouragingly, the HE partition was identical (Figure 4.9 A). We generated an analogous displacement facial signature graph containing exposed individuals in the Caucasian cohort (FAS & HE) producing a similar partition of the HE group into 39 FAS-like and 32 control-like subsets (Figure 4.9B).

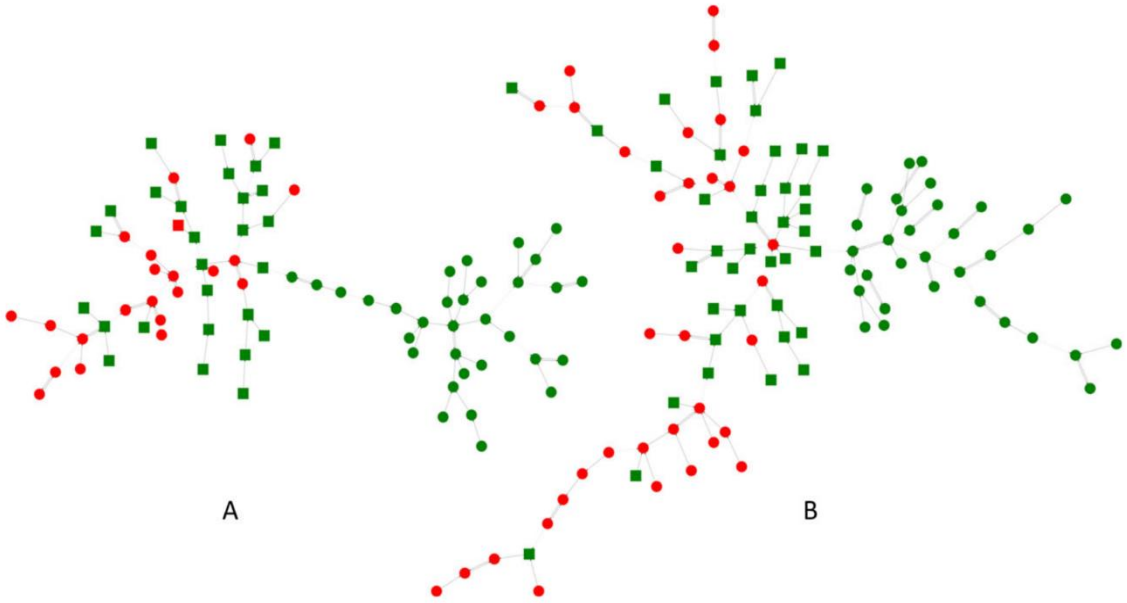


Figure 4.9 Displacement facial signature graphs of FAS (red circles), FAS-like HEs (green squares) and control-like HEs (green circles) for A) Cape Coloured (FAS n=22; HE n=59) and B) Caucasian cohorts (FAS n=35; HE n=71).

We generated normalised displacement signature graphs of Cape Coloured FAS and HE individuals, for midline (nose, upper lip vermilion, philtrum, and midline profile) and non-midline regions (eyes, malar and mandible). In non-midline regions, HEs partitioned into similar FAS-like (HE-FAS) and control-like (HE-HC) subsets as with the full face in previous results, but failed to in midline structure. Refer to Appendix 3 for these graphs where partitioning of HE into FAS-like and control-like subsets is clear for mandible, eyes and malar regions. In contrast, the upper lip vermilion, philtrum, nose and profile do not separate in the same manner.

Cape Coloured individuals partitioned into the HE-HC subgroup of the facial displacement signature graph showed orbital hypertelorism. Figure 4.10 shows the facial signature of horizontal x-axis displacement representing the mean HE-HC subgroup to mean control comparison. Left (red) and right (blue) displacement can be clearly seen in the periorbital regions. We quantify this with significantly increased separation of left-right inner and outer canthi (Figure 4.10 B, D; $p < 0.01$ for both) and an increased interpupillary distance (Figure 4.10 C; $p < 0.01$) calculated with the addition on pupillary landmarks to Cape Coloured HC and HE-HC subsets.

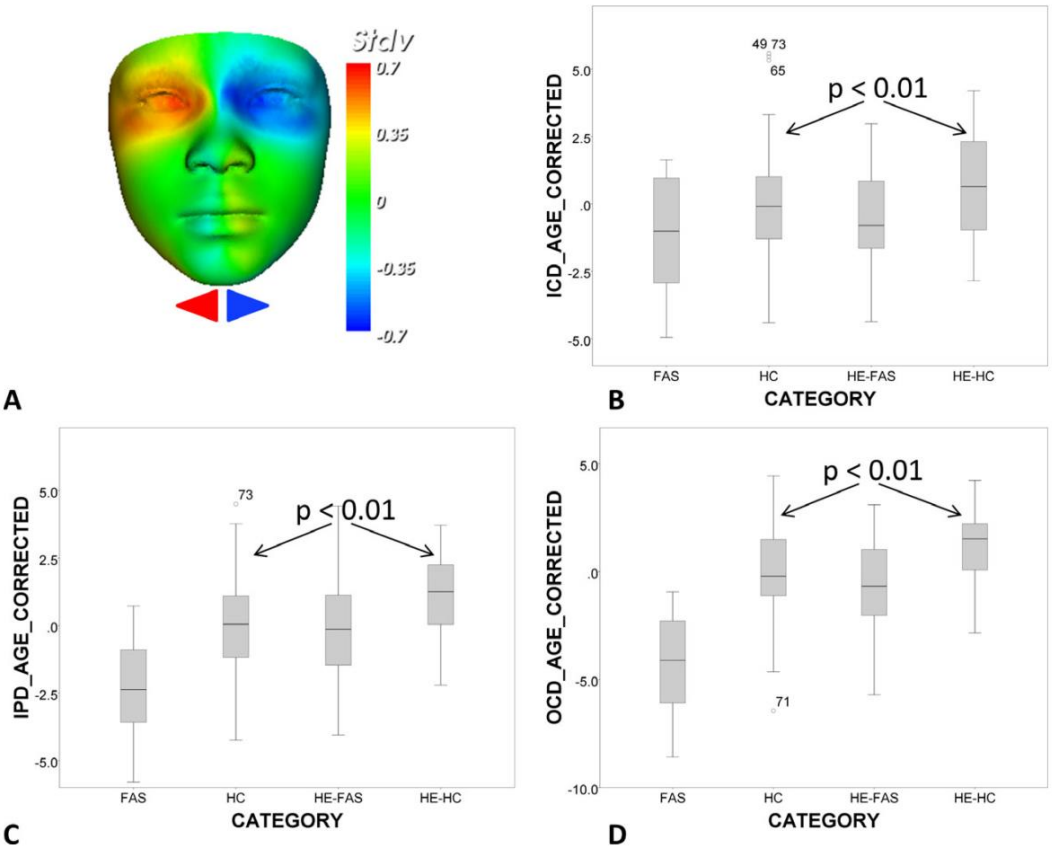


Figure 4.10 Orbital hypertelorism in Cape Coloured control-like HE subgroup (HE-HC). A) Mean Cape Coloured HE-HC to control displacement signature heat map with respect to the horizontal axis. Left (red) and right (blue) surface colouring representing orbital separation at ±0.7SD. Boxplots showing left-right periorbital landmark distances of inner canthi (B), interpupillary distance (C) and outer canthi (D) for FAS, HC, HE-FAS and HE-HC subsets.

4.4.6 Neurobehavioral Differences in the HEs with a Presence or Absence of FAS-like Facial Features

From the analysis in Chapter 3, we found neurobehavioural differences between HEs with FAS-like features and those without. This transpired into a hypothesis suggesting HEs with more FAS-like facial dysmorphism present diminished, more FAS-like cognitive performance. Recapitulating this analysis on the Caucasian signature graph division did not result in the same behaviour. Of the 36 FAS-like Caucasian HEs, 15 had results for CVLT-C and 13 for WISC-IV testing. No significant differences were found between groups across the CVLT-C scores or WISC-IV variables available. Therefore, the suggestion that FAS-like facial dysmorphism in the HE group giving rise to equivalent FAS-like neurocognitive results appears to be limited to the Cape Coloured cohort.

4.5 Conclusions

Clinicians experience difficulty recognising and accurately measuring the FAS-associated facial features, upper-lip vermilion thickness, philtral smoothing and palpebral fissure length essential for diagnosis. Furthermore, prenatally alcohol-exposed individuals who do not express the

typical features, or where dysmorphology is too subtle, pose a more significant clinical challenge. Similar to the aims of the previous chapter, this chapter focuses on developing objective techniques to assist the screening process for identifying facial features associated with PAE. Additionally, in this chapter, we turn our attention to identifying small, subtle differences that may exist between ethnicities across the FASD spectrum.

While FAS is clinically recognisable, we still need to test our techniques on this diagnosis as any screening tool would have test efficacy to discriminate individuals with FAS from controls, given accurately assigned clinical labels. Testing using FAS diagnosed subjects is necessary before we can apply it to other alcohol-exposed subgroups such as the HE subset. In this and the previous chapter, we demonstrated the application of DSMs for the accurate control-FAS discrimination testing of Cape Coloured individuals. However, when applying a similar analysis to localised regions of the face, classification accuracy varied according to region and ethnicity. Inconsistencies between the cohorts were interestingly focused to midline regions of the face, whereas non-midline regions remained consistent in their discriminating power. Potential explanations for this are that PAE could have an increased variability of effects on midline development, or that some ethnic groups have a higher susceptibility to midline effects.

A notable result that for the Control-FAS classification testing showed the malar region in both cohorts to be highly discriminating, almost as accurate as the full face. This result is surprising given how featureless this region is, compared to the mid-face and other tested regions. Measuring malar flattening or midline profile changes by eye in a conventional clinical setting is problematic, if not impossible, without the assistance of imaging analysis. It is certainly less challenging to judge philtral smoothness and thinness of the upper lip-vermilion which are far more convenient markers. What we have shown with these results suggests that with image-based facial analysis, identifying characteristics of other regions of the face may be equally, or even more informative.

The philtrum of the Caucasian cohort provided better Control-FAS discrimination than the Cape Coloured cohort. This was an unexpected result, so to better understand the reasoning for this result we introduced a curvature-based analysis of the face. Before this point, the approach for shape analysis had been to assess localised differences in size normal to the face surface. Given the accurate quantification of philtrum, flatness has a profound impact on the outcome of a FAS diagnosis, it seems natural to consider philtral curvature. Mean surface curvature heat maps provide an intuitive approach to understanding the convexity/concavity of the localised regions of the philtrum. Also, a simple extension to this calculation provides curvature values relative to the surface normal to discriminate curvature in a plane orthogonal to an axis. This axial interpretation allowed curl and groove to be introduced to emphasise directional curvature differences on the philtrum surface.

The philtrum of the Cape Coloured cohort provided poorer Control-FAS discrimination rates compared to the Caucasians, which may partly be explained by the relative philtrum flatness proximal to the nostrils in controls. In contrast, the groove tends to extend the length of the philtrum in Caucasian controls, providing a greater distinction between FAS and control philtrum flatness. Above all, the degree of philtrum groove flattening in the Cape Coloured FAS subgroup appears to be less than in the Caucasian FAS subgroup.

We show homogeneity of the midfacial profile shape difference of Caucasian controls to be higher than that of the Cape Coloured controls. This could be due to the ethnic diversity and

admixture within the Cape Coloured population estimated to be composed of Khoisan (32-43%), Bantu-speaking African (20-36%), European (21-28%), and Asian (9-11%) ancestries (De Wit *et al.*, 2010). Despite this, there was an opposing effect in the in midline profile of the Cape Coloured individuals with a diagnosis of FAS, resulting in greater and more homogeneous dysmorphism. Homogeneity is a likely contributor for improved Control-FAS delineation and may explain the greater discrimination accuracy seen in the Cape Coloured midline profile compared to Caucasians.

Previous studies have suggested a potential overlap of embryonic facial form between FAS and the genetic condition DiGeorge Syndrome (Sulik *et al.*, 1986a; Lipinski *et al.*, 2012), where animal models alcohol-exposed in the equivalent third or fourth week of human gestation have FAS-like or DiGeorge like craniofacial features, respectively. The classic features of DiGeorge syndrome include hypertelorism, so it is encouraging to identify orbital hypertelorism in a subset of the Cape Coloured HE individuals. However, it was not possible to replicate this result in the analogous Caucasian HE partition – instead, we observe only a slight narrowing and upslanting of the periorbital regions.

4.6 Discussion

In this study, we extended the protocol for the analysis of the FASD face from the previous chapter with the addition of a Caucasian cohort, the introduction of curvature analysis and an extended range of facial regions. We focused on facial analysis cohesion between both cohorts and explored inconsistencies brought to light by control-FAS discrimination testing. For the full face, discrimination testing performed consistently well for both cohorts. However, discrimination rates were inconsistent for isolated regions of the face, suggesting that the effect of prenatal alcohol exposure on the facial form may be localised differently and of variable degree between the cohorts. Notably, the midline profile and philtrum regions displayed opposing efficacy in control-FAS discrimination in the two cohorts.

A FASD diagnosis is, of course, based on more than a single facial feature but it is an interesting observation that two cardinal features used for diagnosis, the philtrum and upper-lip vermilion, perform poorly in isolation when discriminating between control and individuals with a FAS diagnosis. Complexities arise when referring to the diagnostic criteria used for FAS diagnosis, which permits the absence of 1 of the three typical facial features. Lip and philtrum chart scores of 4 or 5 denote a thin upper lip and smooth philtrum, and for the Cape Coloured cohort was assigned to 17/22 (77%) and 16/22 (73%), respectively. Analogous lip and philtrum proportions for the Caucasian FAS subgroup were similar with 27/35 (77%) and 30/35 (86%), respectively. The “1 of 3” criteria rule creates facial heterogeneity within the FAS subgroups and partly explains why shape-based discrimination rates (Figure 4.7 and Table 4.2) are lower than those of the full face and other regions.

Additional to the variability of typical FAS features, there is concern over their recognition. Different clinicians diagnosed the two cohorts considered here with no overlap between cohorts. FAS facial features and growth deficits were examined independently by expert dysmorphologists blinded to prenatal alcohol exposure information. A standard protocol for diagnosis (Hoyme *et al.*, 2005) employed the Astley Lip-Philtrum Guide for lip vermilion and philtrum assessment, with palpebral fissure length measured manually. There was a definitive agreement between clinicians on the measurements. The Caucasian cohort underwent a similar

clinical examination, using the same diagnostic protocol, but with a single expert dysmorphologist with extensive experience in FAS. It is important to note, the reliability of the lip-philtrum guide is necessarily subjective and provides only a categorical scale open to misinterpretation and error.

Only a handful of studies compare craniofacial morphology in FASD across different ethnicities. In this chapter, as with the previous, we have considered individuals with confirmed prenatal alcohol exposure who do not meet criteria for a FAS or PFAS diagnosis, whereas previous studies have not included a group similar to our HE category. Additionally, previous studies did not employ surface shape analysis or utilise facial curvature. One such study (Moore *et al.*, 2007), evaluated ethnic differences in facial features in 276 individuals recruited from 4 ethnic cohorts: Cape Coloured (n=103), Finnish Caucasian (n=99), African American (n=24), and North American Caucasian (n=50). Landmark annotations manually placed on 3D images were used to derive sixteen anthropometric measures, and for each cohort, discriminant analysis was performed to identify the best combination of age and the facial measurements to support control-FAS discrimination testing. Each cohort had a unique, although often overlapping, set of variables used the final classification model, with at least one ocular-related measurement, reduced PFL and reduced inner or outer canthal distance. Control-FAS discrimination rates for each of the four ethnic groups were 92%, 93%, 79% and 77% respectively. May and colleagues (2010) produced the largest and most comprehensive study of the ethnic difference in dysmorphology associated with prenatal alcohol exposure. Unlike the previous study, both FAS and PFAS diagnoses were considered across three ethnic populations: Plains Indians (n= 321), Italians (n = 300), and South Africans (n = 300). The major finding from this study concluded that approximately 20% of the variability in dysmorphology of core features were associated with ethnicity. Multilevel model techniques were employed finding variables that predicted dysmorphology across all three cohorts to be shortened PFL, smooth philtrum, flat nasal bridge, and fifth finger clinodactyly. Amongst the findings, the South African controls were found to have higher percentages of thin vermilion, smooth philtrum, flat nasal bridge, shortened interpupillary distance and anteverted nostrils. This is consistent with our results for the Cape Coloured philtrum analysis, but also we were able to use curvature analysis to clarify these differences in greater detail.

It is important to address specific limitations that exist within our study when comparing to those previously published. Firstly, while we had access to an additional African American cohort, unfortunately, the sample size of alcohol-exposed individuals was insufficient to produce reliable results. This will hopefully be addressed in future studies. 3D image acquisition for this study was carried out using a series of near-identical commercial 3D stereo-photogrammetry systems deployed across sites. Previous comparisons have shown near perfect image parity between such camera systems (Tzou *et al.*, 2014). Cohort sizes are relatively small, in particular regarding individuals with a FAS diagnosis. While age-ethnicity matched sets used for signature normalisation contain the same number of controls, the age distribution in both cohorts is non-uniform (Figure 4.5). There are also gaps in the continuity of age, particularly in the Cape Coloured cohort, which will affect the matched means. Additionally, small sample sizes have restricted our ability to address effects of BMI and gender on the facial form. We have shown how ethnicity-specific changes in underlying facial morphology affect the detection of features associated with prenatal alcohol exposure.

In a study investigating the nutritional and health status of 2 disadvantaged communities in the Western Cape Province (Oelofse *et al.*, 2002), stunting and underweight were more prevalent in coloured infants (18% and 7%, respectively) than in black infants (8% and 2%, respectively). Prevalence of anaemia was high for both coloured (64%) and black infants (83%), and zinc deficiency was prevalent in 35% and 33% respectively. In a longitudinal study of a Cape Coloured population, length/weight and head circumference were measured at 6.5 and 12 months and 5 and 9 years (Colin Carter *et al.*, 2012). Prenatal alcohol alcohol-related reductions in weight, length/height and head circumference were found to be exacerbated by iron deficiency anaemia among infants but were not modified at 5yrs. It has been concluded that in Cape Coloured children, the effect of compromised nutrition is a major contributor to reduced weight, head circumference and palpebral fissure length which remain important conjectures of FASD (May *et al.*, 2010).

In a recent study of a Cape Town FASD cohort, poor maternal nutrition affected both the alcohol-exposed and controls groups (Carter *et al.*, 2016), but no evidence was found to suggest a relationship between prenatal alcohol exposure and stunting. Due to the similarity of in utero nutritional status between groups, poor nutrition cannot account for the observed effect of prenatal alcohol exposure associated facial features. However, animal studies have observed that differences in nutrition and caloric intake in utero alter the effects of alcohol exposure (Shankar et al., 2006; Weinberg et al., 1990). In this chapter, we observed notable differences in prenatal alcohol associated facial morphology between the Cape Coloured and Caucasian cohorts in the midline and non-midline regions. A few studies have addressed the maternal nutritional status of FASD populations in from South African communities (May *et al.*, 2014, 2016). In the earlier of these studies, undernourishment was reported in all of the participating mothers of both control and FASD groups, but the severity of malnourishment worsened in the FASD group. Compared to controls, FASD mothers reported a significantly lower intake of calcium, docosapentaenoic acid, riboflavin, and choline, all key nutrients important for growth and development and therefore capable of disrupting face and brain development. In this publication, it is proposed that poor nutrition may lead to an increase in the severity of FASD features. In the latter study (2016), all mothers reported undernourishment, but mothers of FASD children had an increased dietary intake of 13 of 25 micro-nutrients compared to mothers of controls. Although there is a suggestion that prenatal malnutrition may cause deficits of growth and neurodevelopment, there appears to be no human or animal study demonstrating the influence of nutritional status on dysmorphology. The two prior mentioned studies (May *et al.*, 2014, 2016) recorded current diets of women with school-aged children with FASD to make implications about their nutritional intake during pregnancy. In another prospective study of a South African Cape Coloured population (Carter *et al.*, 2016) assessing diet and anthropometric measures of heavy drinking pregnant women, they found similar results; both controls and heavy drinkers reported poor gestational weight gain and a deficiency of key dietary nutrients. No significant associations were found between maternal diet or anthropometry and maternal alcohol intake during pregnancy. Thus, we conclude that the influences of prenatal nutrition are unlikely to confound or mediate the results presented in this chapter.

In summary, we have shown in this chapter how facial curvature can assist the evaluation of facial features of prenatal alcohol exposure and also utilised this to explain inconsistencies of control-FAS discrimination rates between ethnic cohorts. In particular, we have shown how facial curvature can help to identify definite differences. Additionally, we identified a subset of

heavy prenatally exposed individuals exhibiting significant orbital hypertelorism, often associated with DiGeorge syndrome. This result is interesting given the long-standing suggestion that prenatal alcohol exposure at a particular gestational time point, for example, week 4, can contribute to an increased hemispheric separation of the brain with a subsequent increase in orbital separation (Parnell et al., 2009; Sulik et al., 1986).

5 Modelling Combined Face-Brain Morphology in Foetal Alcohol Spectrum Disorders

In this chapter, we introduce a DSM based approach for the multi-structural modelling of 3D morphology. We utilise this technique to investigate the effects of prenatal alcohol exposure on midline and non-midline brain structure morphology and observe relationships with facial features. DSM techniques are extended to produce a multimodal analysis of face-brain dysmorphology by employing combined dense surface models.

5.1 Introduction

The previous two chapters of this thesis and preceding studies (Suttie *et al.*, 2013, 2017) focused on improving the understanding of facial dysmorphism across the FASD spectrum, and have been goal orientated to assist clinicians in the recognition of associated facial features. Using DSM analysis of facial form we have successfully applied discrimination testing to accurately discriminate between FAS and control subgroups. In this chapter, we introduce brain morphology into our analysis, and introduce novel DSM techniques for combined face-brain analysis. We utilise an existing Caucasian cohort recruited within the CIFASD consortium which partially forms a subset of the dataset used in Chapter 4. Here, we pair seven localised regions of the face (previously referred to in Chapter 4; illustrated in Figure 4.1) with two brain regions known to be disproportionately affected by prenatal alcohol exposure (PAE); the corpus callosum (CC) and the caudate nucleus (CN).

Facial and brain dysmorphia associated with FASD have been well characterised independently in the literature. Only a handful of studies to date have investigated correlations between them (Roussotte, 2012; Yang, Phillips, *et al.*, 2012). Notably, previous work utilising DSMs (Lipinski *et al.*, 2012) introduced the simultaneous analysis of face and brain in a murine model of *in utero* ethanol-exposure at two time points (gestational days 7 and 8.5) and concluded that stage-specific ethanol exposure caused a unique pattern of face-brain dysmorphism. While human studies can never provide alcohol exposure data with such precision, it remains important to understand face-brain correlations in alcohol-exposed populations. In this study, we introduce and apply combined DSM techniques for face-brain shape analysis.

This analysis utilises an existing dataset comprising high resolution 3D facial images acquired from a 3D stereophotogrammetry; and segmented structural MRI images from 3T scanners. Diagnostic categories in this chapter are analogous to those in the previous chapter. As we explain in Chapter 4, the partial foetal alcohol syndrome (PFAS) label is excluded from this cohort, leaving three categories; foetal alcohol syndrome (FAS), non-syndromal heavily exposed (HE) and non-exposed controls. We previously identified differences that exist between midline and non-midline facial regions between ethnic cohorts. Consequently, brain regions selected for this study are representative of both midline (corpus callosum) and non-midline (caudate *nucleus*) neurostructural components.

5.2 Study Population

The dataset used for this study consists of 120 participants forming a partial subset of the Caucasian cohort from the previous chapter, due to the scarcity of image data available we also include Hispanic individuals. Throughout its existence, the CIFASD consortium has recruited several thousand participants obtaining clinical examination data, neurobehavioral profiles, 3D images and MRI scans. For each participant in this analysis, we require clinical data, a 3D facial image and an MRI image acquired at time point close to that of the face. CIFASD recruited participants from four sites across the USA; San Diego (n=48), Los Angeles (n=37), Atlanta (n=8) and Minneapolis (n=27). Self-reported ethnicity data collected for all participants revealed 28 individuals to be of Latin American descent, while the remaining are of European ancestry. Diagnostic categorization splits the dataset into the following classifications: control (n=48); FAS (n=22); HE (n=54). We also refer to further superset categorisation simply as ‘exposed’ which encompasses both HE and FAS groups. The interval between MRI and 3D face data acquisition is on average 3 months (SD 4.5 months) with those with a period of greater than 9 months between scans excluded.

Clinical assessment

Subjects with a history of brain injury, genetic disorders and non-alcohol related neurocognitive deficits were excluded. Each child was examined for growth deficits and screened for FAS facial features by a multiple specialised dysmorphologists utilising standard protocols for measurements and diagnostic labelling (Hoyme *et al.*, 2005). Upper lip vermilion and philtrum assessment employed the Astley and Clarren Lip-Philtrum Guide (Astley and Clarren, 2000), and palpebral fissure length (PFL) was measured using a manually positioned ruler. Confirmation of excessive alcohol consumption comparable to alcohol abuse patterns was recorded for all subjects within the ‘exposed’ group. Given the small number of available images, controls with reported ‘minimal’ alcohol exposure were also included in this study (n=3; where minimal

exposure ≤ 1 drink per week (US) and never exceeding 2 drinks per occasion). Refer to Table 3.2 in Chapter 3 for criteria used to differentiate FAS, PFAS and non-syndromal HE individuals.

As part of the clinical assessment, clinicians administered standardised batteries of neuropsychological tests which were available for 70 of the 120 subjects. Individuals in each category (controls (n=26); HE (n=30); FAS (n=14)) were assessed using the Differential Ability Scales, Second Edition (DAS-2; Elliott 2007) and the California Verbal Learning Tests for Children (CVLT-C; Delis et al. 1994). The DAS-2 consists of a series of 20 cognitive subtests measuring verbal, nonverbal, and spatial reasoning abilities. In addition, DAS-2 offers a measure of general cognitive ability (GCA) with outcomes similar to an intelligence quotient (IQ) score. The CVLT-C consists of a comprehensive series of tests designed to assess verbal learning, recall and memory.

5.3 Methods

5.3.1 MRI Image Pre-processing

Before the application of combined DSM, images from both modalities require a degree of pre-processing. We derive a 2D corpus callosum representations from a midline contour tracing of points from the raw MRI scans. With the addition of connected equidistant left-right points the 2D contour transforms to a 3D surface providing a ~6mm midsagittal representation. Caudate nucleus segmentation was computed using an automated multi-atlas based segmentation method (NiftySeg) for which the marching cubes algorithm was used for the extraction of a 3D surface representation.

Corpus Callosum Extraction

For the CC extraction, each subject was registered to MNI152 space via FSL's flirt (Jenkinson *et al.*, 2012). A midline slice is selected, and the trace starts from the posterior aspect of the genu to the inferior-most point of the splenium for both the top and bottom segments of the CC. The points for the upper and lower segments are then concatenated to form a continuous midline CC contour representation. Equidistant +/- Z points are generated at a set distance either side of each vertex; edges are drawn from these points to the original contour points to create an artificially triangulated surface mesh, surface normals are then calculated by taking the vector cross product of the created edges. An unweighted sliding-average smoothing algorithm was applied to eliminate noise from the mesh. Landmarks were applied to the 3D surface using an in-house software tool to add three reliable points manually; the lower point of the splenium, the tip of the genu and a midpoint along the upper arch. These three points are used to generate a clipping plane through the midline of the 3D surface. Surface geometry is extracted from the intersecting points within this plane to produce an accurate representation of a midline contour. The resulting contour is used to produce 30 additional equidistant semi-landmarks between the three original points; these are calculated based on equal geodesic distance and determined by the shortest path along the contour. The resulting semi-landmarks (Figure 5.1C) consist of 10 from the genu to the mid-upper point, 10 from the genu to the upper and 10 on the inferior arch from the genu to the splenium.

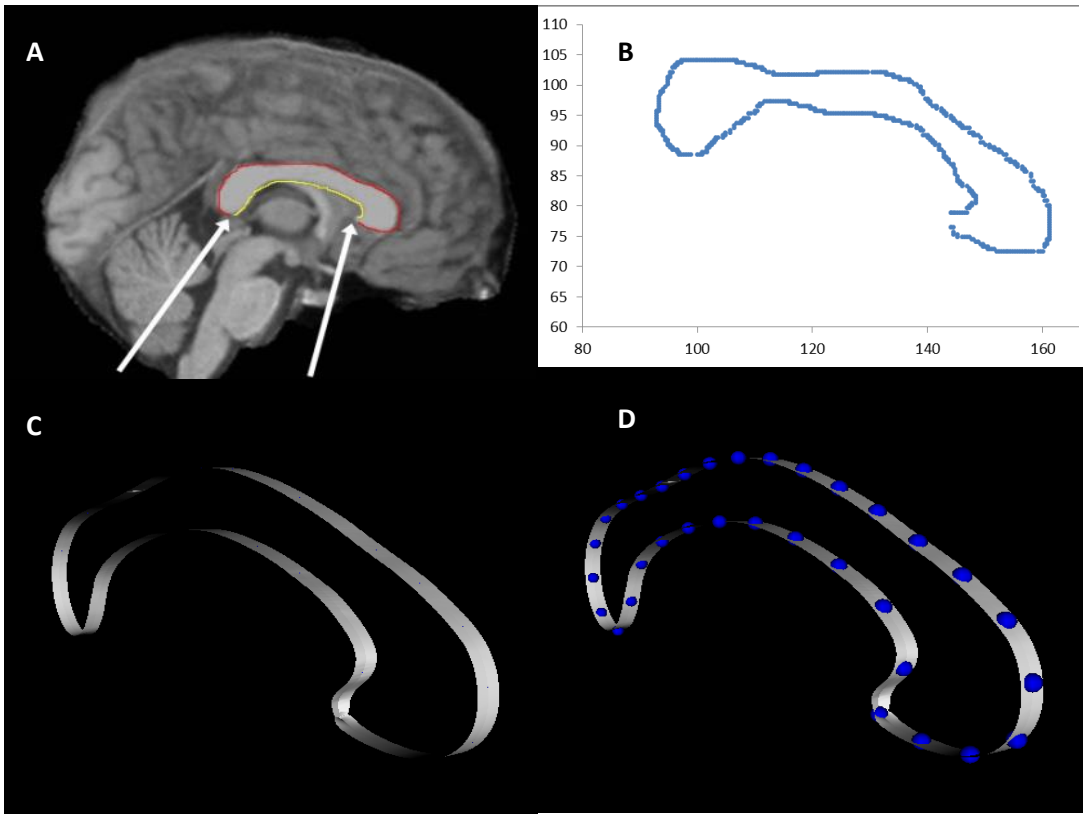


Figure 5.1 (A) Original contour outline on MRI, splenium and genu points (B) Plot of raw CC points in 2D from contour outline (C) 3D rendered and laplacian smoothed CC from raw points using additional equidistant vertices on x-axis (D) 33 landmarks including 35 semi-automatically induced equidistant semi-landmark points.

Caudate Nuclei Extraction

Raw MRI images were segmented using an automated multi-atlas pipeline developed by researchers from the *Translational Imaging Group* (*TIG*) at University College London. 20 control MRI atlas images containing 67 manually discriminated regions were obtained from the Biomedical Image Analysis Group at Imperial College London (www.brain-development.org). Employing an open source image processing tool; NiftyReg (Ourselin *et al.*, 2000; Modat *et al.*, 2010) we apply an initial image affine registration and non-rigid registration prior to segmentation. Using the multi-STEPS (Multi-label Similarity and Truth Estimation) algorithm packages into NiftySeg (Jorge Cardoso *et al.*, 2013) a parameter optimised, multi-atlas based propagation process calculated accurate parcellations for each subject. Thresholding the image as per the original atlas images provided the caudate nucleus segmentation. The selected region contains the extracted body of the nuclei, omitting the inter-connected tail and putamen regions. MicroView (http://microview.parallax-innovations.com), a freely available 3D image viewer and analysis platform, was used to extract a 3D surface representation of the CN. Using in-house software, 6 anatomical landmarks were manually annotated to the 3D surface for each left and right section of the CN.

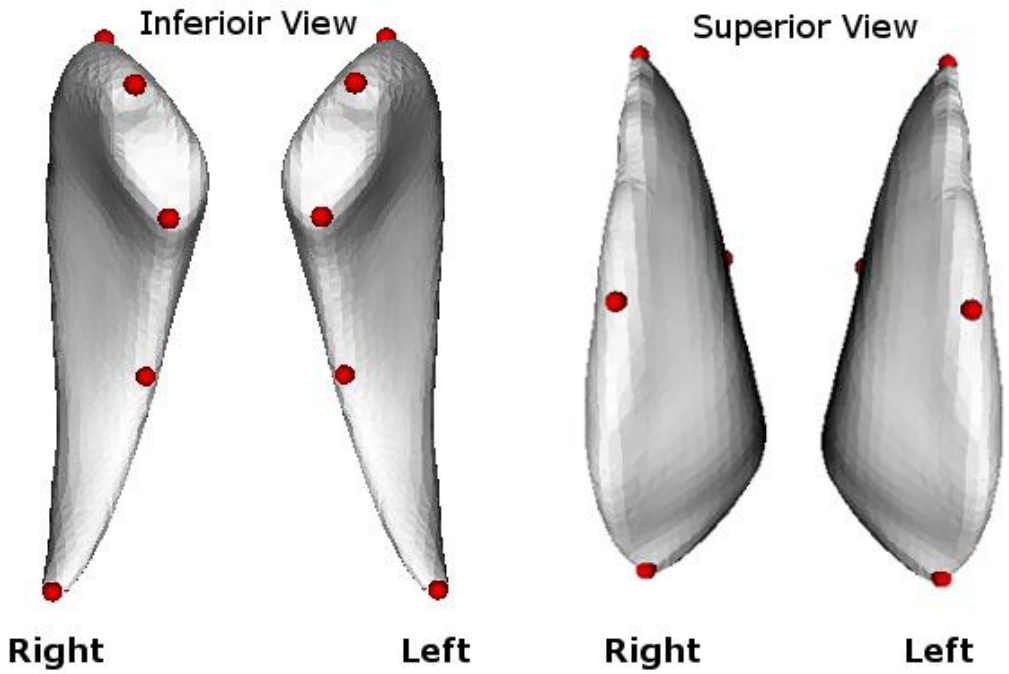


Figure 5.2 Landmark placement for the caudate nucleus. Three landmarks are placed on the head, two on superior and inferior midpoints of the body and one on the tail of the left and right caudate nucleus.

5.3.2 Combining DSM Models

Combined dense surface models (CDSMs) exploit morphological covariance that exists between any two anatomical structures given a 3D surface representation. 3D facial images obtained by conventional off the shelf hardware such as stereo-photogrammetry cameras can be combined with surfaces generated from MRI images of brain components to produce a multi-structural model. To achieve this, we utilise a construct based on Active Appearance Models (Cootes *et al.*, 1998) where shape models of 2D structures were able to take into account combined representations of shape and intensity. Likewise, this technique extends the methodology of textured DSMs (Hutton 2004) where PCA models of shape and appearance are combined. Textured DSM face models assimilate the inputs of two separate models, the 3D face shape model and a model encompassing the mapped pixel value from each vertex obtained from 2D photos. Combining the models involves rebuilding a DSM with the concatenated inputs of both models, the result being a photorealistic shape model (Figure 5.3). As lighting effects can influence PC values within these models diagnostic classification testing and shape analysis is unreliable. Instead, a key motivation for these photorealistic models was their potential role in training clinicians to identify facially affected conditions. While following the same theoretical approach, PC values calculated from combined DSM models of separate anatomical components, such as face-brain components, produce PC values that express the shape variance of both structures (Figure 5.4). To build a combined DSM we first generate independent sets of PCA values from each shape model and then combine these models into a single DSM by concatenating the input vectors.

However, in a typical scenario such as face and brain modelling, the different structures in each model may vastly differ in size and shape variation causing disproportionately weighted

eigenvalues. To overcome this issue, models must be made commensurate with a scaling factor. This is achieved in a similar way to that in Hutton 2004, which utilised a simple method using the ratio of total variances to scale one model to match to the other.

In previous face only DSM studies in FASDs (Suttie et al. 2017; Suttie et al. 2013) we utilise DSM shape models with PC values reduced to explain 99% of shape variation. However, when DSMs that are to be combined in this manner, it makes sense to utilise all modes of variation since each mode of a has the potential to correlate with any mode. Thus, each DSM model resynthesizes individual 3D surfaces as a weighted linear sum of PCs that account for 100% of the shape variation and will produce inputs of n-1 PC values. Each example or *pair* is now represented by a vector of concatenated eigenvalues.

Figure 5.3 Facial synthesis at the first 3 PCs in a Bardet-Biedl syndrome textured DSM model (n=60; age range: 6-66yrs) represented at -3 (left), 0 (centre) and +3 (right) standard deviations. PC1 value expresses mainly growth; PC2 indicates a significant variance in skin tone combined with broadening of the face; PC3 predominately representing an elongation of the face with subtle surface-texture changes.

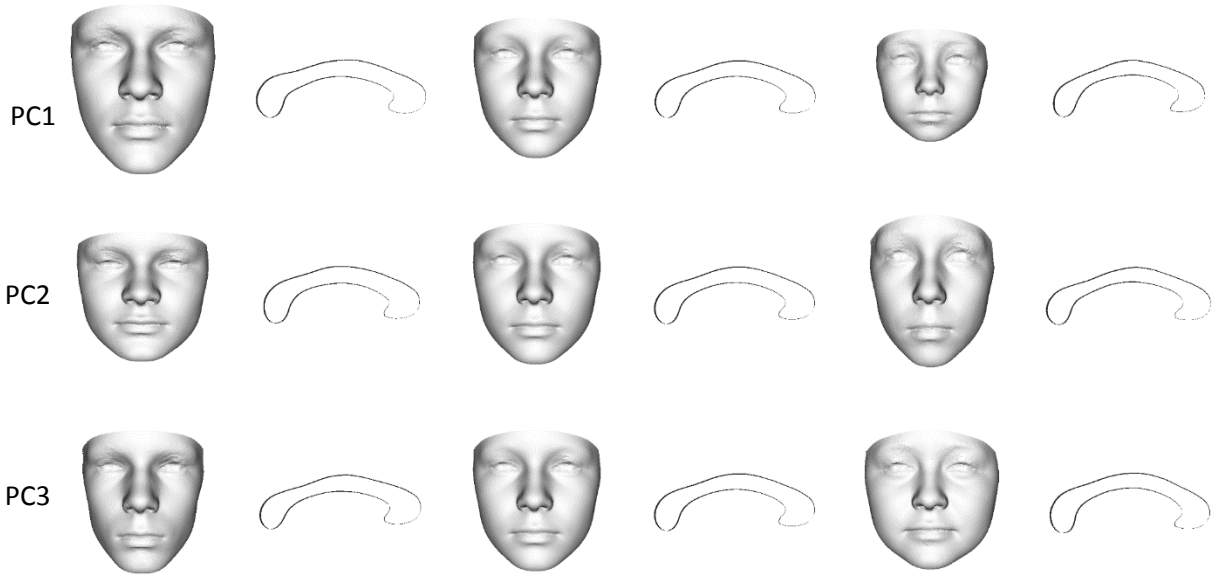


Figure 5.4 Face-brain synthesis at the first 3 PCs in a control-FASD model (n=125) represented at -3 (left), 0 (centre) and +3 (right) standard deviations. PC1 value expresses mainly growth in the face with thickness differences in the CC; PC2 indicates facial elongation with the CC showing regional callosal thickness differences; PC3 predominately representing depth changes in the face and a variance in the radius of the CC arch.

5.4 Results

5.4.1 Cohort Characteristics

The male-female split for the 120 subjects was balanced without age distribution differences for FAS and control groups. However, the HE group has a significant majority of male subjects (72%). Across the sample, we observed the presence of dose-dependent effects for head circumference, FAS < HE & control; for height and weight, FAS < HE & control; and for volumetric brain measures (left and right caudate volume and total intracranial volume) FAS & HE < control. Sample characteristics are detailed in Table 5.1.

	FAS (n=22)	HE (n=50)	Control (n=48)	F
Age at 3D photo (yrs)	13.0 ± 3.2	12.3 ± 2.2	12.6 ± 2.3	0.84
Age at MRI image (yrs)	1.6 ± 3.2	12.2 ± 2.0	12.6 ± 2.3	1.26
3D-MRI difference (mths)	2.0 ± 2.1	3.2 ± 4.6	4.2 ± 4.7	1.90
Sex n (% male)	10 (45.5%)	36 (72.0%)	26 (45.5%)	2.17
Height (cm)	136.7 ± 18	147.1 ± 18	154.1 ± 14.8	8.18***
Weight (kg)	33.7 ± 11.6	45 ± 19	50.0 ± 15.0	7.47***
Head circumference (cm)	51 ± 1.8	53.6 ± 2.5	54.9 ± 2.0	22.49***
BMI	17.3 ± 2.2	19.9 ± 4.4	20.6 ± 4.0	5.34**
Facial anomalies				
Short palpebral fissures	19 (86.4%)	8 (16%)	4 (8.5%)	39.73***
Flat philtrum	12 (54.5%)	11 (22%)	9 (19.1%)	6.19**
Thin vermillion	12 (54.5%)	14 (28%)	11 (23.4%)	3.13*
Brain measures				
Left caudate volume	3307 ± 669	3714 ± 566	4048 ± 471	13.27***
Right caudate volume	3361 ± 693	3800 ± 577	4713 ± 499	10.59***
Total cranial volume	1173365±230611	1267130±280067	1373410±235361	4.09*
Child's FSIQ (SD)	-0.61 ± 0.8	-0.50 ± 0.9	0.57 ± 0.9	21.71***

Table 5.1. Sample characteristics; values represent mean ± SD or %, F statistic produced from one-way ANOVA test. Dose-dependent relations for head circumference: FAS < HE and controls ($p < .001$), height and weight: FAS < HE and controls ($p < .03$), BMI FAS < HE and controls ($p < .01$), palpebral fissure length FAS < HE and controls ($p < .0001$), flat philtrum and thin vermillion FAS < HE and controls ($p < 0.05$) and FSIQ FAS and HE < controls ($p < .0006$). Dose dependent relations for brain measures: left and right caudate volume FAS and HE < controls ($p < .001$), total intracranial volume FAS < control ($p < .002$) and HE < control ($p = 0.03$). There were no significant left-right caudate volume differences for any of the 3 groups either controlled by gender or mixed.

* $P < 0.05$.

** $P < 0.01$.

*** $P < 0.001$.

5.4.2 FAS-Control Classification of Face, Corpus Callosum and Caudate Nucleus Shape

Similar to the discrimination testing carried out previously in Chapters 2 and 3, we apply multi-folded discrimination testing for the full face and smaller localised regions, using 20 randomly sampled splits to determine 90% to 10% training and unseen test sets respectively. Additionally, we apply discrimination testing to the brain regions; corpus callosum and caudate nucleus (additionally split into left-right segments) using the same method. We tested agreement of control-FAS discrimination based only on surface shape using CM, SVM and LDA algorithms achieving expectedly high overall discrimination rates in concordance with clinical diagnoses for the midline facial regions (profile, philtrum, nose and upper-lip vermilions) and non-midline regions (face, eyes, mandible and malar). Similarly, results for the brain regions performed well, with the corpus callosum and caudate nucleus (whole and divided into left and right sides) discriminating as accurately as the facial regions.

	CM	**LDA**	**SVM**
Face			
Full face	0.93	0.93	0.95
Profile	0.83	0.88	0.90
Eyes	0.90	0.93	0.93
Malar	0.95	0.95	0.93
Mandible	0.88	0.88	0.90
Nose	0.95	0.95	0.98
Lip vermilion	0.85	0.85	0.85
Philtrum	0.83	0.83	0.88
Brain			
Corpus callosum	0.90	0.90	0.93
Caudate nucleus	0.88	0.88	0.88
Left caudate nucleus	0.83	0.85	0.90
Right caudate nucleus	0.83	0.85	0.90

Table 5.2 Separate face and brain discrimination rates estimated as the mean area under the ROC curves of 20 cross-validation tests, and correspond to the probability of accurately classifying pairs of individuals; 1 taken from the control set and the other from the exposed set. Testing for the full face and localised facial regions revealed high discrimination rates as did the brain components: corpus callosum and caudate nucleus.

5.4.3 FAS-Control Classification Using Combined Face-Brain Morphology

We applied control-FAS discrimination testing, analogous to the methods used for the separate face and brain evaluation, to combined models of face and brain. In this manner, training and

test sets are constructed from the PCs of combined DSM models encompassing the full face, and the seven localised facial regions in combination with the surface of either the corpus callosum or the caudate nucleus. When combining face and corpus callosum, we observed equal or marginally improved results compared to face and corpus callosum separately. To simplify the results table, and to show where improvements in accuracy occur, Figure 5.5 and Figure 5.6 plot combined face-brain vs face only discrimination results for each facial region. Notably, face-brain combinations show the greatest increase over face-only discrimination rates when combining corpus callosum with midline facial structures; nose, lip vermillion, profile and philtrum (Figure 5.5A). Combining caudate nucleus with the face increased classification accuracy for the eyes and whole face (Figure 5.5B) compared to face only. However, improved discrimination rates occurred when combining left caudate nucleus with facial regions (Figure 5.6).

	+Corpus callosum			**+Caudate nucleus**			**+left caudate nucleus**			**+right caudate nucleus**		
	CM	LDA	SVM	CM	LDA	SVM	CM	LDA	SVM	CM	LDA	SVM
Full face	0.90	0.90	0.93	0.90	0.90	0.90	0.95	0.95	0.95	0.90	0.90	0.93
Profile	0.90	0.90	0.88	0.85	0.85	0.88	0.95	0.95	0.95	0.90	0.90	0.90
Eyes	0.90	0.93	0.93	0.90	0.90	0.90	0.95	0.95	0.98	0.95	0.95	0.95
Malar	0.90	0.93	0.93	0.90	0.90	0.93	0.98	0.95	0.95	0.93	0.93	0.93
Mandible	0.88	0.88	0.90	0.88	0.88	0.88	0.95	0.95	0.95	0.90	0.90	0.93
Nose	1.0	1.0	1.0	0.93	0.93	0.88	0.95	0.95	0.95	0.93	0.93	0.93
Lip vermilion	0.90	0.90	0.90	0.85	0.85	0.93	0.95	0.95	0.95	0.85	0.85	0.88
Philtrum	0.95	0.98	0.98	0.85	0.85	0.88	0.95	0.95	0.95	0.90	0.90	0.88

Table 5.3 Combined face-brain discrimination rates estimated as the mean area under the ROC curves of 20 cross-validation tests, and correspond to the probability of accurately classifying pairs of individuals; 1 taken from the control set and the other from the FAS set. Testing for the full face and localised facial regions combined with corpus callosum, whole caudate nucleus and left right caudate nucleus revealed high discrimination rates and improved accuracy compared to analogous testing methods on face and brain components separately.

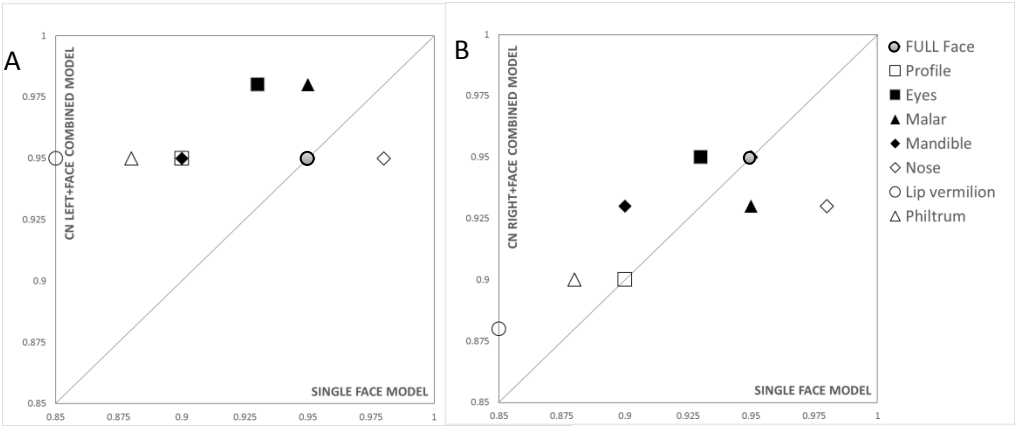


Figure 5.5 Discrimination rates for face only (X-axis) vs combined face-brain (Y-axis) for A) left and B) right caudate nucleus. Midline facial structures (nose, lip vermillion, profile and philtrum) are labelled with unfilled symbols, and non-midline structures (full face, eyes, malar and mandible) represented by filled symbols. Only moderate improvements were observed when combining right caudate nucleus with facial regions with the exception of nose and malar (B). In contrast, left caudate nucleus improved discrimination rates when combined with lip vermillion (0.95), malar (0.98), nose (0.95), eyes (0.98) and profile (0.95). Face-brain classification for left or right caudate nucleus did not differ substantially for midline and non-midline facial regions.

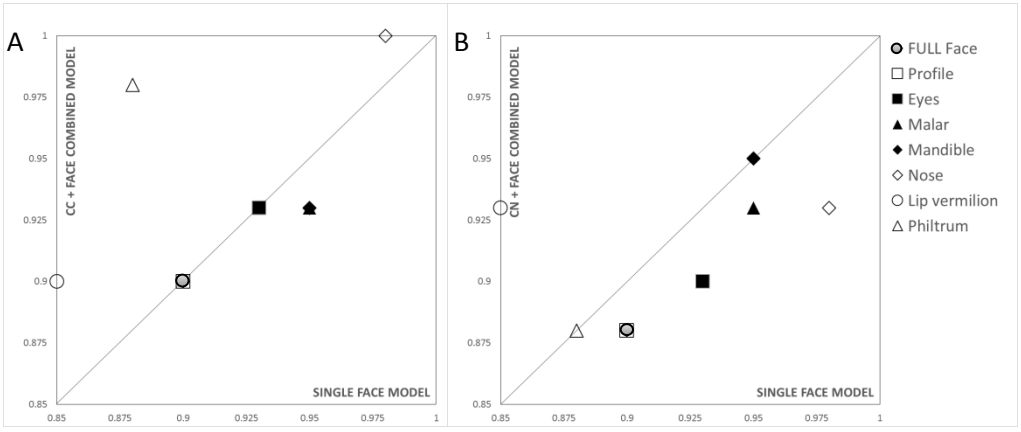


Figure 5.6 Discrimination rates for face only (X-axis) vs face-brain (Y-axis) for A) corpus callosum and B) the caudate nucleus (whole). Midline facial structures (nose, lip vermillion, profile and philtrum) are labelled with unfilled symbols, and non-midline structures (full face, eyes, malar and mandible) by filled symbols. For the corpus callosum, face-only classification accuracy is increased in combination with midline facial regions (nose (1.0), lip vermillion (0.90), philtrum (0.98)) but unchanged for profile (0.90). Adding the entire caudate nucleus improved classification for lip vermillion (0.90) but not mandible (0.90) and philtrum (0.98). All other face regions (eyes, mandible, profile, malar and nose) gave inferior classification results when combined with brain components.

5.4.4 Face-Brain Growth DSM Representation

In each of the combined models calculated (face+corpus callosum, face+caudate nucleus), the first mode of variation (PC1) continues to be representative of growth as in previous DSM face models (Hammond & Suttie 2012; Chapters 3 & 4) and the face models in this cohort (Figure 5.8). However, in the single corpus callosum model, neither PC1 nor any other PC values are associated with age for controls or alcohol-exposed groups (Figure 5.7), thus, the corpus callosum does not share the same associations as the age-related growth trajectories of the face. PC1 in the caudate nucleus model has a strong correlation with caudate volume ($r^2 = 0.75$) for all categories within the model, and shows a subtle growth trajectory with age (Figure 5.9).

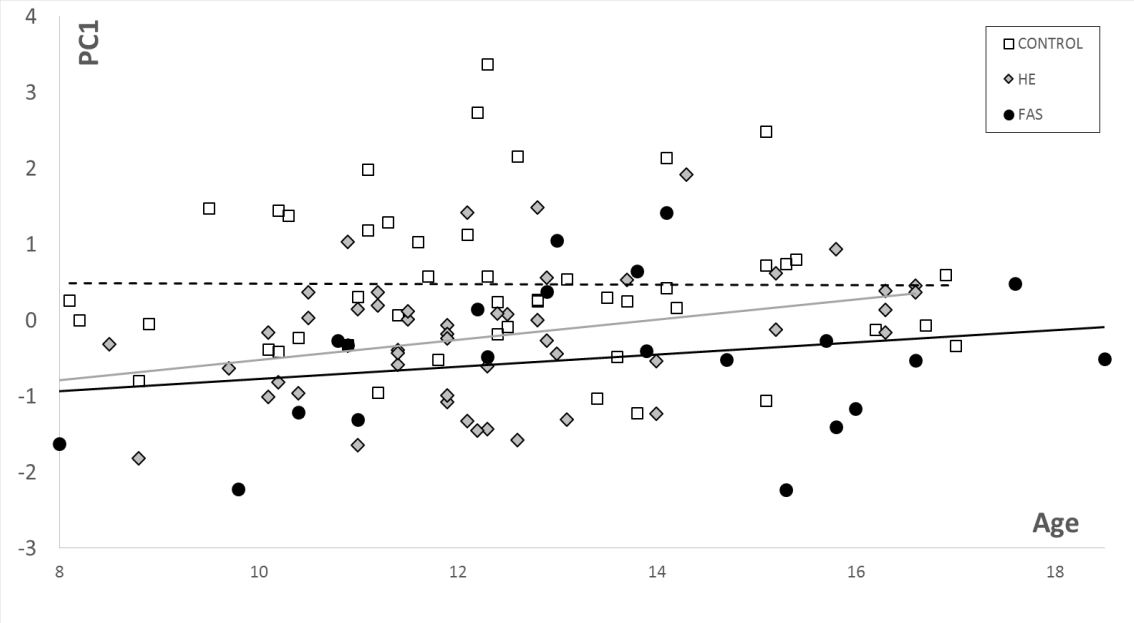


Figure 5.7 Growth comparison with controls using PC1 from a single shape model of the corpus callosum.

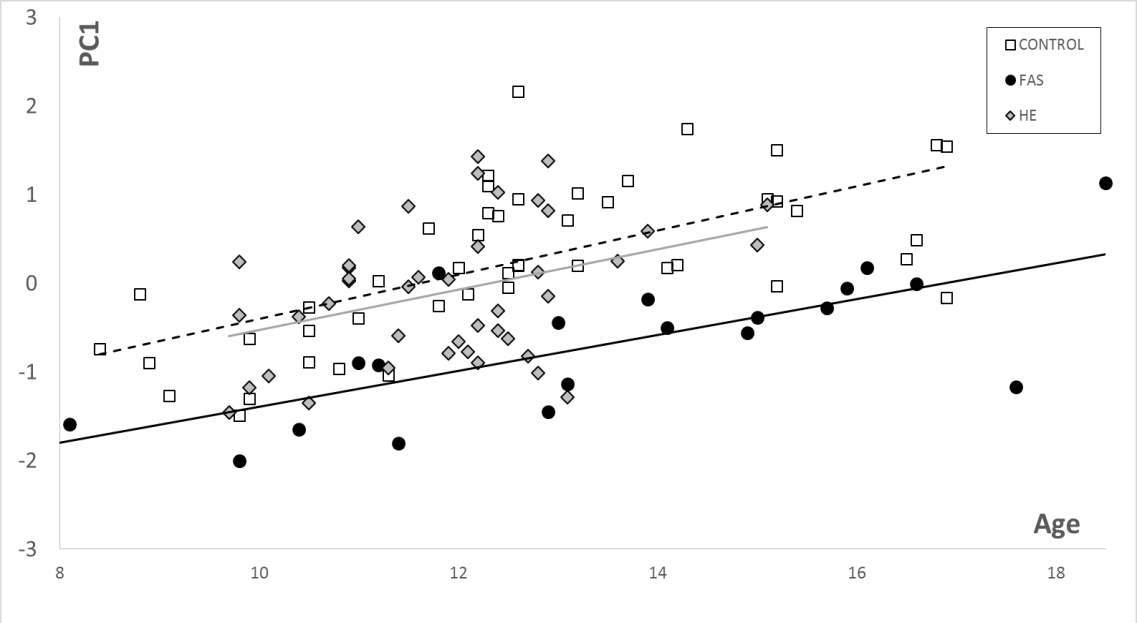


Figure 5.8 Growth comparison with controls using PC1 from a single shape model of the face, showing significantly lower growth for FAS ($p < 0.0001$) and marginally lower for HEs ($p < 0.06$).

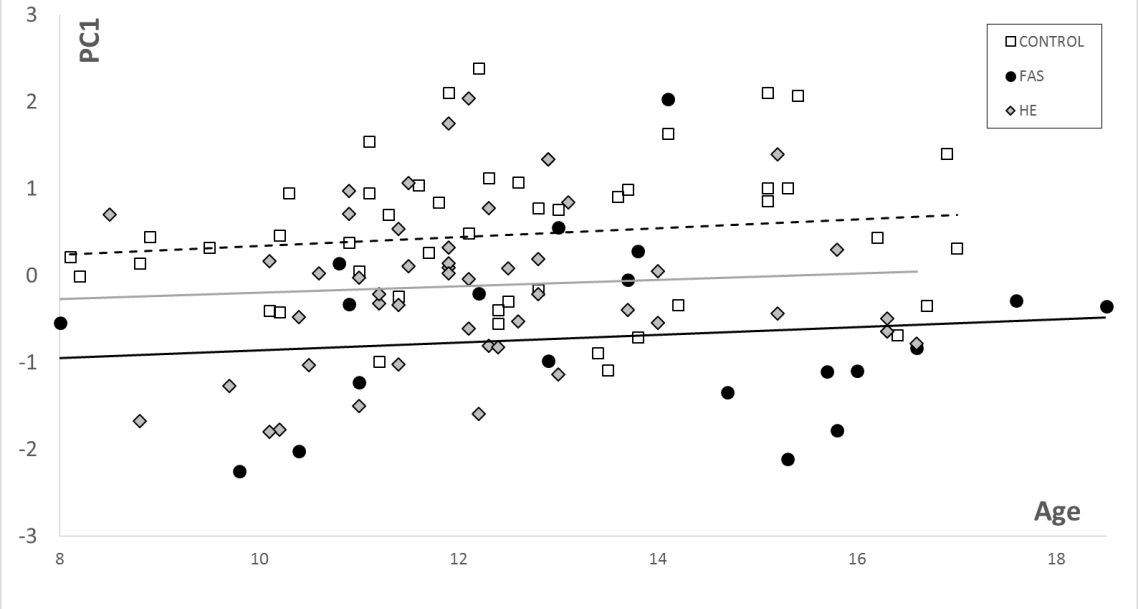


Figure 5.9 Growth comparison with controls using PC1 from a single shape model of the caudate nucleus, showing significantly lower growth for FAS ($p < 0.00001$) and HE ($p < 0.002$).

5.4.5 Signature Graphs of the Caudate Nucleus, and Face, Divide Heavily Exposed (HE) Individuals

In the previous two chapters, signature graphs containing exposed individuals (FAS, PFAS & HEs) partitioned HEs into FAS-like and control-like subgroups based on facial dysmorphology. For this analysis we apply the same methodology to this dataset to provide graphs of exposed individuals using both brain components and the full face. Generating an equivalent displacement facial

signature graph on this cohort, produced a comparable partition of the HE group into FAS-like and control-like subsets (figure supplementary) to those seen in previous analyses (Chapters 3 & 2; Suttie et al, 2013, 2016). Similarly, with a graph constructed of displacement differences of the caudate nucleus (Figure 5.10A,B) a distinctive HE divide partitions HEs into 2 subsets containing 17 subjects with an affinity to the FAS cluster with 29 remaining. This division is justified by the edges of the upper nodes on subgraph 8, where connections to the HE dominant subgraphs 9 and 11 are only by secondary links. By secondary, we refer to the linking of two signature clusters by the two closest signatures in the different groups (Figure 5.10C).

Utilising the caudate nucleus partitioned FAS-like subset, we generated mean facial signatures (Figure 5.11B) which show notable similarity to the FAS mean signatures (Figure 5.112A) across each axial comparison. When applying the same signature graph analysis to the corpus callosum, we fail to identify any clustering of HE individuals amongst FAS nodes. Instead, both HE and FAS nodes were dispersed almost evenly across the graph (see Appendix 5) with dispersion indices for FAS (0.76±0.06) and HE (0.56±0.04) indicating a lack of homogeneity. HE individuals that have an affinity to clusters of FAS nodes had similar dysmorphism for face and brain. We generated mean signatures of the face from the extracted FAS+PFAS like and control-like HEs from the CN graph. Using the calculated mean we applied closest means testing of the facial model to identify the position relative to control and FAS+PFAS means Additionally, we plot the faces of all HE individuals in this FAS+PFAS cluster to determine their relative position to control and FAS calculated means (Figure 5.13). With the exception of the eyes, this caudate nucleus derived FAS-like mean is closest to FAS means rather than control means for each facial region.

MRI is often subject to imaging disparities between machines which could influence our results given the multiple sites used for data collection. Control numbers for each site group (Los Angeles, Atlanta, Minnesota and San Diego) are not sufficient for generating age-site matched means. Instead, to test the construction of the caudate signature graph we reconstruct the graph labelling each node by site rather than diagnosis. Using a collapsed form of this site-labelled graph (see Appendix 5), we calculate an entropy-like measure, dispersion index, to summarise clustering association or dispersion of members of each site (Hammond et al. 2012; Appendix 2). Site dispersion coefficients for caudate nucleus of the exposed subset with a CI 95% for each site are SD=0.77±0.03, MN=0.80±0.04, LA=0.79±0.06, suggesting there is no site based affinity in this caudate signature graph.

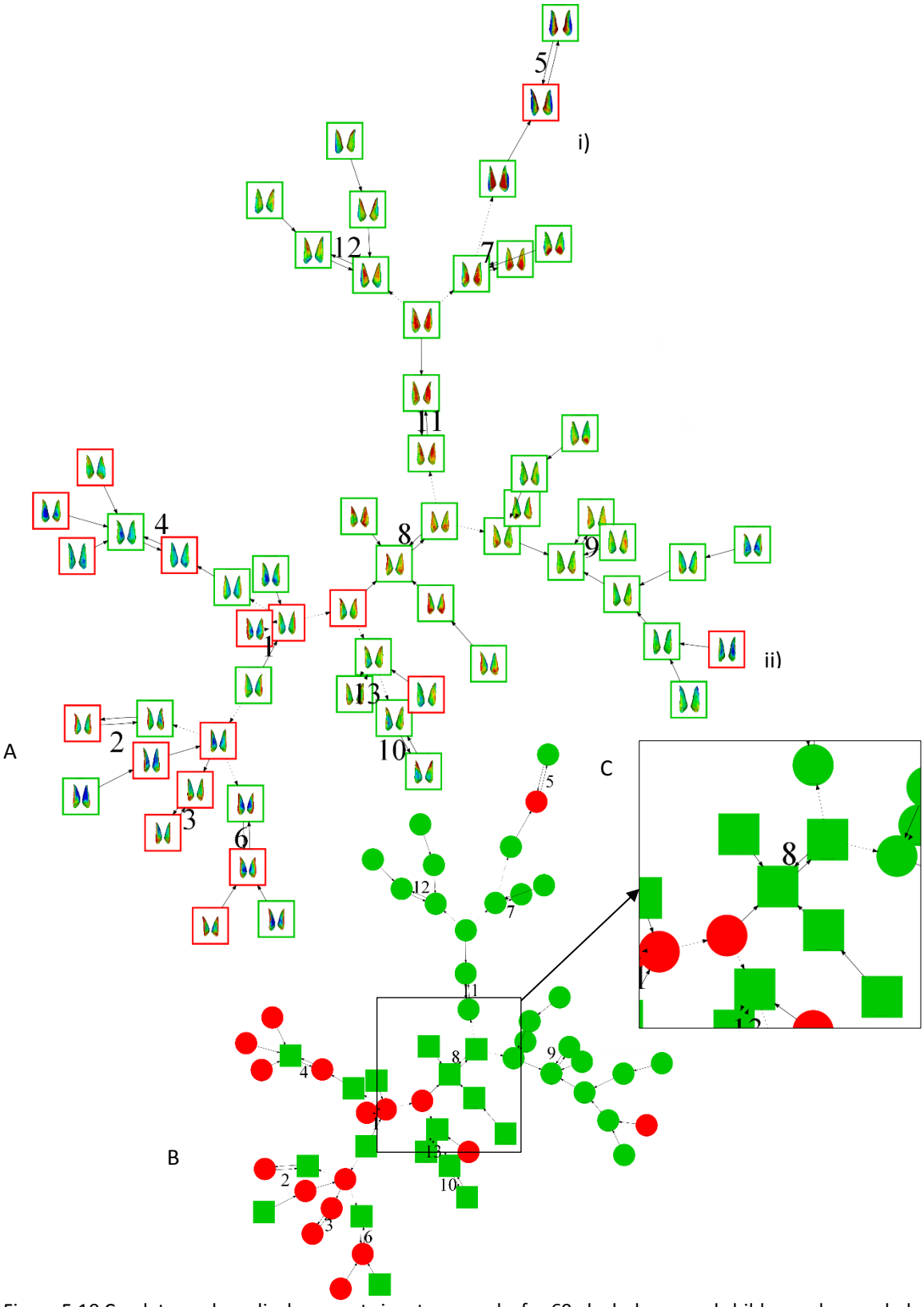


Figure 5.10 Caudate nucleus displacement signature graphs for 69 alcohol-exposed children colour coded as FAS (red, n=17) and HE (green, n=42). A) Graph is shown with caudate nucleus signatures and similar to previous facial analysis (Chapters 3 & 4) there is a FAS-like clustering in a subset of HE individuals which divides the graph under subgraph 8. A.i & ii, show 2 FAS outliers that do not cluster with the other individuals with a FAS diagnosis. B) Colour coded binary representation with FAS-like HE nodes (green squared, n=17) and the remaining HEs (green circles, n=29) to illustrate the upper and lower graph separation. C) Zoomed view of division at subgraph 8, where the last node is connected to subgraphs 11 and 9 by only secondary connections.

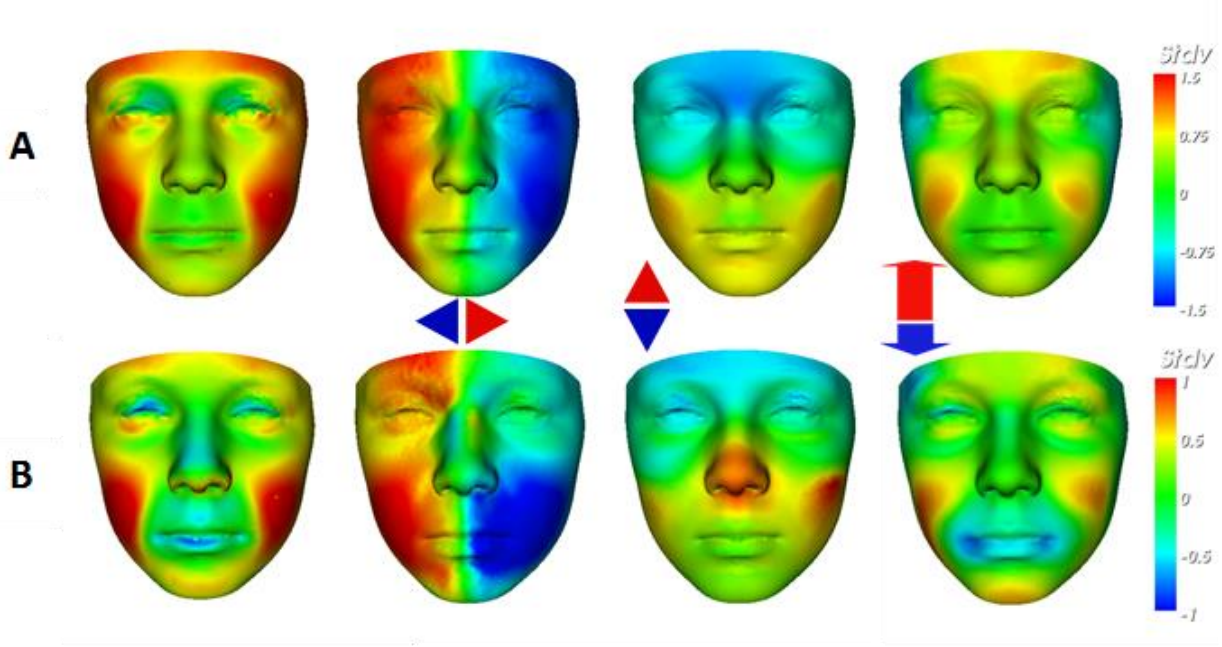


Figure 5.12 Comparing mean facial signatures along each axis between A) FAS mean (n=22) at ±1.5SD and B) the FAS-like HE partition (n=17) at ±1SD induced in the caudate nucleus signature graph. This visualisation allows us to identify regions of similar dysmorphism; malar flattening (red in mid-face; cols 1 & 4); philtrum smoothness (faint blue on philtrum groove; col 1); retrognathia (red on the chin; col 4); overall size reduction (red in col 1 and lateral red-blue in col 2). The FAS-like partition on the horizontal axis (B, col 3) also shows shortening of nose (red on columella) which is a FAS associated feature but is not visible in the equivalent FAS mean signature (A, col 3).

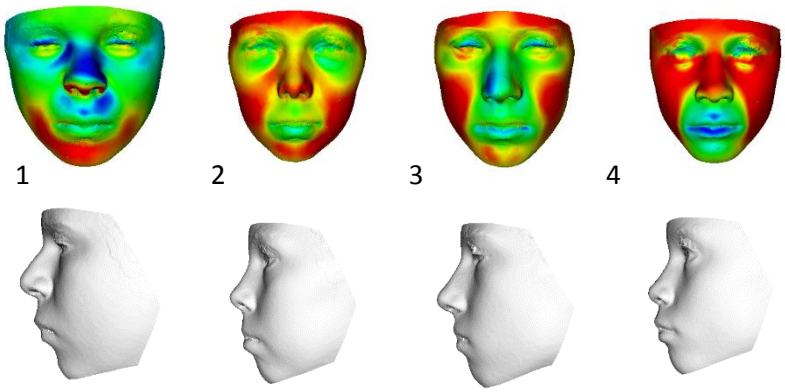


Figure 5.11 5 facial signatures from the FAS-like HE partition generated from the caudate nucleus signature graph using 25 age-matched controls. Retrognathia can be identified from the first 4 of these signatures (red on the chin) and the last 4 show an indication of malar/midfacial flattening (red midface) and overall reduction in head size (red on the forebrain region). Signatures 1 and 5 show philtral smoothness (blue on the philtrum centre).

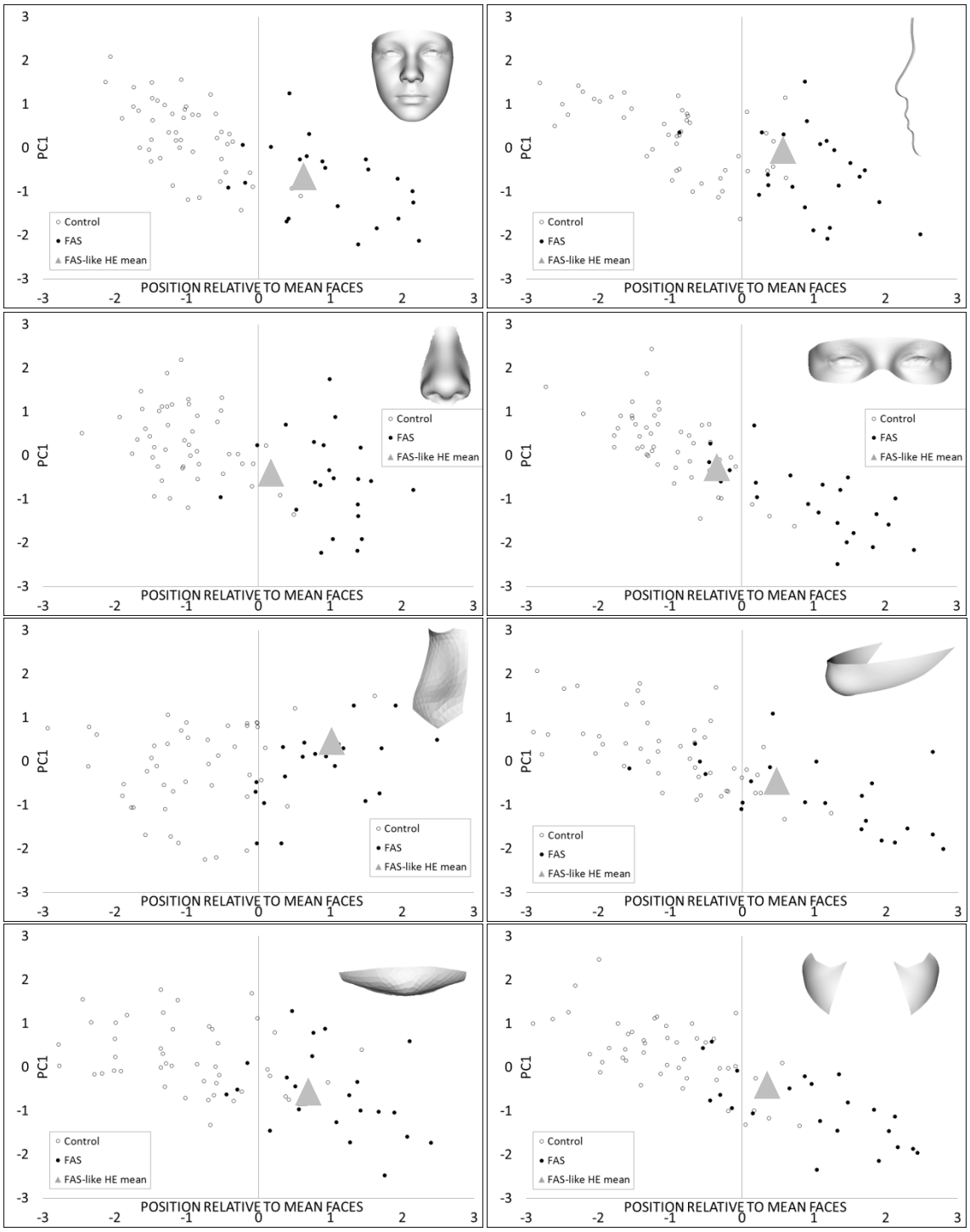


Figure 5.13 Classifying the FAS-like HE mean derived from the caudate nucleus HE partition for each facial region. The FAS-like HE mean for each region; face, profile, nose, eyes, philtrum, mandible, upper lip vermillion, and malar, is plotted on a hyperline between the control mean (-1) and the FAS mean (+1) against PC1. In each case, with the exception of the eyes, the FAS-like HE derived from the caudate nucleus signature partition is closest to the FAS mean.

5.4.6 Face-Brain Shape Correlations

Using combined face-brain models, we observe correlative relationships within exposed subsets. From the discrimination testing, it is evident that specific face regions combined with the corpus callosum and the caudate nucleus produce accurate control-FAS discrimination rates (Table 5.3). In particular, combining the corpus callosum with the philtrum, nose, profile and lip-vermillion yielded improved results when compared with single models of each.

Additionally, we generate animations showing dynamic morphs for a more intuitive visualisation of these face-brain correlations (http://www.ucl.ac.uk/~sejjmfj/face_brain_morphs.htm). These videos show dynamic morphs between FAS and control means for combined face-brain models of the corpus callosum plus (i) the philtrum, (ii) nose, (iii) profile and (iv) lip vermillion.

5.4.7 Caudate Nucleus Left-Right Asymmetry

In the FAS-control classification results stated in section 5.4.3 and shown in Table 5.3, we see improved discrimination rates combining the left caudate with facial regions, compared to the right caudate. This suggests that asymmetry of the caudate nucleus may play a role in FAS-control delineation. Asymmetry of a biological component is typically analysed by comparing original and reflected forms (Klingenberg et al. 2002; Hennessy et al. 2005). Similar to previous DSM asymmetry analyses (Hammond *et al.*, 2008; Postema *et al.*, 2017; Schoot *et al.*, 2017), we generated a reflected form of each caudate surface and swapped the left-right landmarks before resampling the surface, effectively generating a dense correspondence of surface points on both left and right sides of the caudate. To consider asymmetry, we constructed a DSM of the combined set of original caudate surfaces and their reflected forms and computed as a simple measure (asymmetry index) as the generalised Euclidean distance between the DSM-based vectors representing each caudate and its reflected form. Illustrated in Figure 5.14 we observe a significantly increased left-right caudate nucleus asymmetry in the control group compared to FAS (Figure 5.14 $p<0.002$). However, when observing the HE group (Figure 5.14 B) the mean asymmetry indexes show no differences compared to controls. Colour-coded mean original to mean reflected comparisons (Figure 5.15) visualise asymmetry as a distance parallel to two orthogonal axes, anteroposterior (depth) and craniocaudal (vertical).

To ensure no site-wide differences are influencing asymmetry indices, we test means from each site by their respective clinical categorisation and for all subjects. There are no significant differences for mean asymmetry indices between sites across the entire dataset or for any clinical category (Table 5.4).

	ATL	LA	MN	SD	F	P-value
Control	13.13 ± 1.18	12.33 ± 4.02	12.09 ± 5.2	11.61 ± 1.1	2.82	0.69
FAS	9.69 ± 0.23	10.58 ± 1.50	9.22 ± 1.84	11.11 ± 3.46	3.16	0.25
HE	9.88 ± 0.35	11.51 ± 3.72	11.02 ± 2.99	12.31 ± 3.82	2.80	0.10
ALL	10.6 ± 1.65	11.69 ± 1.93	11.06 ± 1.57	11.97 ± 2.08	2.07	0.11

Table 5.4 Table S1 Comparing mean asymmetry indices between sites, for each diagnostic category and for all subjects. One-way ANOVA analysis shows no significant differences exist across sites. ATL=Atlanta, LA=Los Angeles, MN=Minnesota and SD=San Diego.

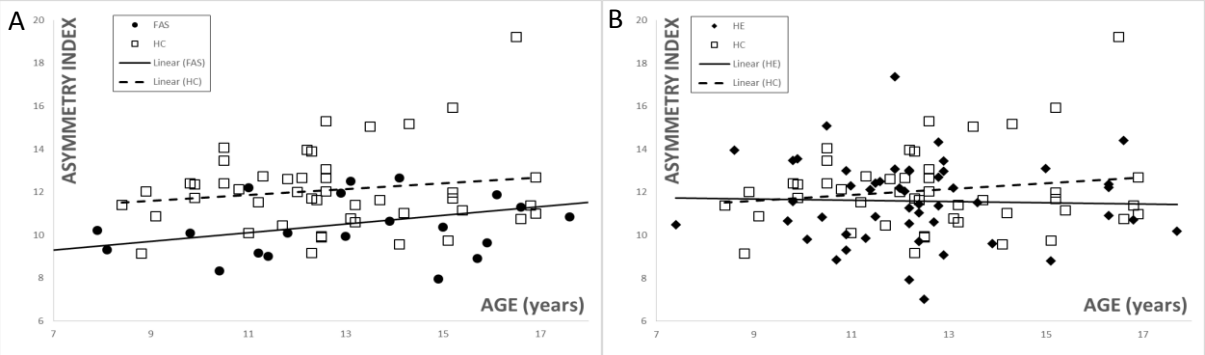


Figure 5.14 Caudate nucleus asymmetry index (distance between whole caudate nucleus and its reflection) against age for A) FAS vs Control and B) HE vs Control. The control group has a significantly higher mean index (12.1) level of asymmetry compared to the mean index (10.5) of the FAS group ($p < 0.002$), however when comparing controls and HE means we see no significant differences.

Localised regions of left-right asymmetry can be visualised by heat mapping displacement differences between the mean reflected surfaces for each group against the originals (Fig 9). From the mean original-reflected differences we observe localised axial asymmetries along the anteroposterior (depth) and craniocaudal (vertical) axes. The depth axis control heat map shows localised directional displacement on the inferior part of the caudate. This localised asymmetry is also evident in the FAS group but to a lesser degree, but is absent in the HEs. A vertical axis comparison shows control growth in a localised region of the superior part of the left caudate compared to the right which is absent in exposed groups. While the mean asymmetry index of HEs is comparable with controls, we do not see the same distinctive axial asymmetry indicating this difference is heterogeneous and not one-sided.

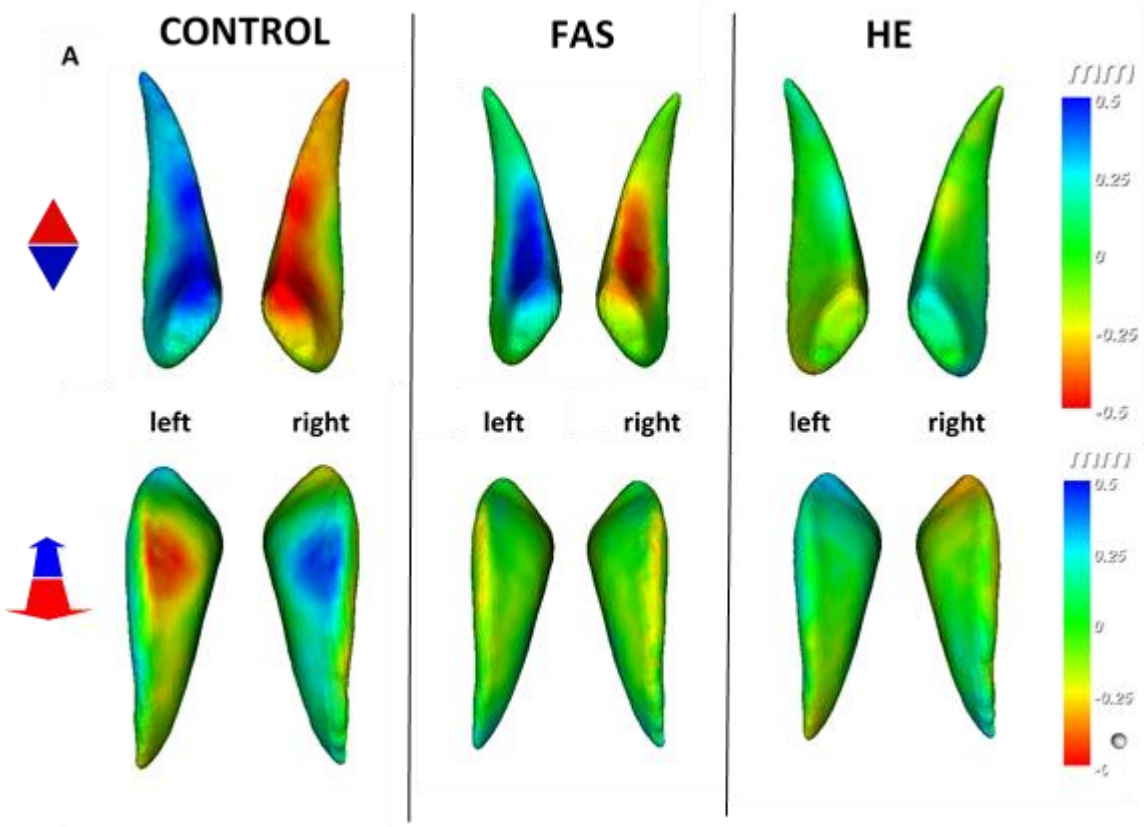


Figure 5.15 Mean axial asymmetry represented by calculating displacement of mean original to mean reflected caudate nucleus for control, FAS and HE groups along A) anteroposterior (depth) and B) craniocaudal (vertical) axes ±0.5mm. A) Shows indication of a left anterior dominance in controls and similarly in FAS but to a lesser extent and localised. The HE mean at the same scale shows no left-right surface differences in that region. B) On a vertical axes we see localised dominance on the superior part of the left caudate, but FAS and HE means show no evidence of the same asymmetry.

5.4.8 Caudate Asymmetry Correlations with Neurocognitive Performance in Prenatal Alcohol Exposure

The control population appear to express greater degrees of directional and overall asymmetry compared to the PAE groups. Testing PC values of a model built only from individuals with PAE (FAS n=22; HE n=54) we observe numerous correlations with neurocognitive test scores on PC4. This PC is primarily responsible for left-right caudate asymmetry, but testing PCs against neurocognitive performance is not a robust method for analysis given that they represent non-specific shape variation in a model. Instead, we utilise the asymmetry index described in 5.4.7 and find several significant positive correlations between CVLT-C and DAS-2 scores and the level of caudate asymmetry in the exposed population. We tested this relationship by computing 2-tailed Pearson correlation coefficients for correlations between the caudate asymmetry indices, and each of the age-adjusted test outcomes of the CVLT-C and DAS-2 batteries. Results indicated significant correlations between asymmetry indices and age-adjusted z-scores for CVLT-C tests: discriminability (r^2=0.13, p<0.02); Long delay cued recall (r^2=0.11, p<0.03); short delay cued recall (r^2=0.13, p<0.03); and a near significant correlation for correct recognition hits (r^2=0.08, p=0.065). Testing DAS-2 standardised scores and caudate asymmetry index revealed the same patterns of significant correlations for: general cognitive ability (GCA) (r^2=0.1, p<0.04); spatial ability (r^2=0.09, p<0.05); and, ability scores: recall of designs (r^2=0.1, p<0.04); sequential and

quantitative reasoning ability score (r^2=0.11, p<0.03); and the verbal similarities ability score (r^2=0.13, p<0.02). In contrast, there were no significant or near significant correlations in the control group. The neurocognitive-caudate asymmetry correlation appears to have an opposing effect compared to that of the alcohol exposed population. Figure 5.16 illustrates these significant correlations, and sample characteristics for neurocognitive tests can be seen in Table 5.5. Details of test outcome variables used are available in Appendix 4.

Test	FAS (n=14)	HE (n=30)	CONTROL(n=26)	F
CVLT-C				
Long Delay Cued Recall	-0.73 ± 0.97	-0.18 ± 0.63	0.13 ± 0.63	3.29 *
Short Delay Cued Recall	-0.9 ± 0.6	-0.25 ± 0.9	0.13 ±0.7	5.08 **
Discriminability	-0.62 ± 0.9	-0.4 ± 1.3	0.25 ± 0.5	3.69 *
Correct Recognition	-0.23 ± 0.7	0.0 ±0.7	0.2 ± 0.5	1.34
DAS-2 (standard scores)				
Spatial Ability	91.9 ± 45.3	91.4 ± 46.0	109.1 ± 49.0	13.6 ***
General Cognitive Ability	90.5 ± 45.1	91.4 ± 45.9	109.8 ± 49.7	12.9 ***
DAS-2 (ability scores)				
Sequential and quantitative	125 ± 69	121 ± 77	155 ± 74	8.3 ***
Verbal Similarities	110 ± 56	115 ± 64	131 ± 65	7.8 ***
Recall of Designs	93 ± 47	95 ± 55	113 ± 54	13.0 ***

Table 5.5 Sample characteristics of CVLT-C and DAS-2 tests which showed correlations with caudate asymmetry index. Values represent the mean ± SD for each outcome variable. F statistic calculated from one-way ANOVA analysis. Dose-dependent relations shown for Long Delay Cued Recall; FAS < control (p < .001); Short Delay Cued Recall; FAS < control (p < .001); Discriminability: FAS & HE < control (0.03); Spatial ability: FAS & HE < control (p < .0001); GCA = FAS & HE < control (p < .001); Verbal Similarities FAS & HE < control (p < .0002); and, Recall of Designs FAS & HE < control (p < .0002).

* P< 0.05.

** P< 0.005.

*** P< 0.0005.

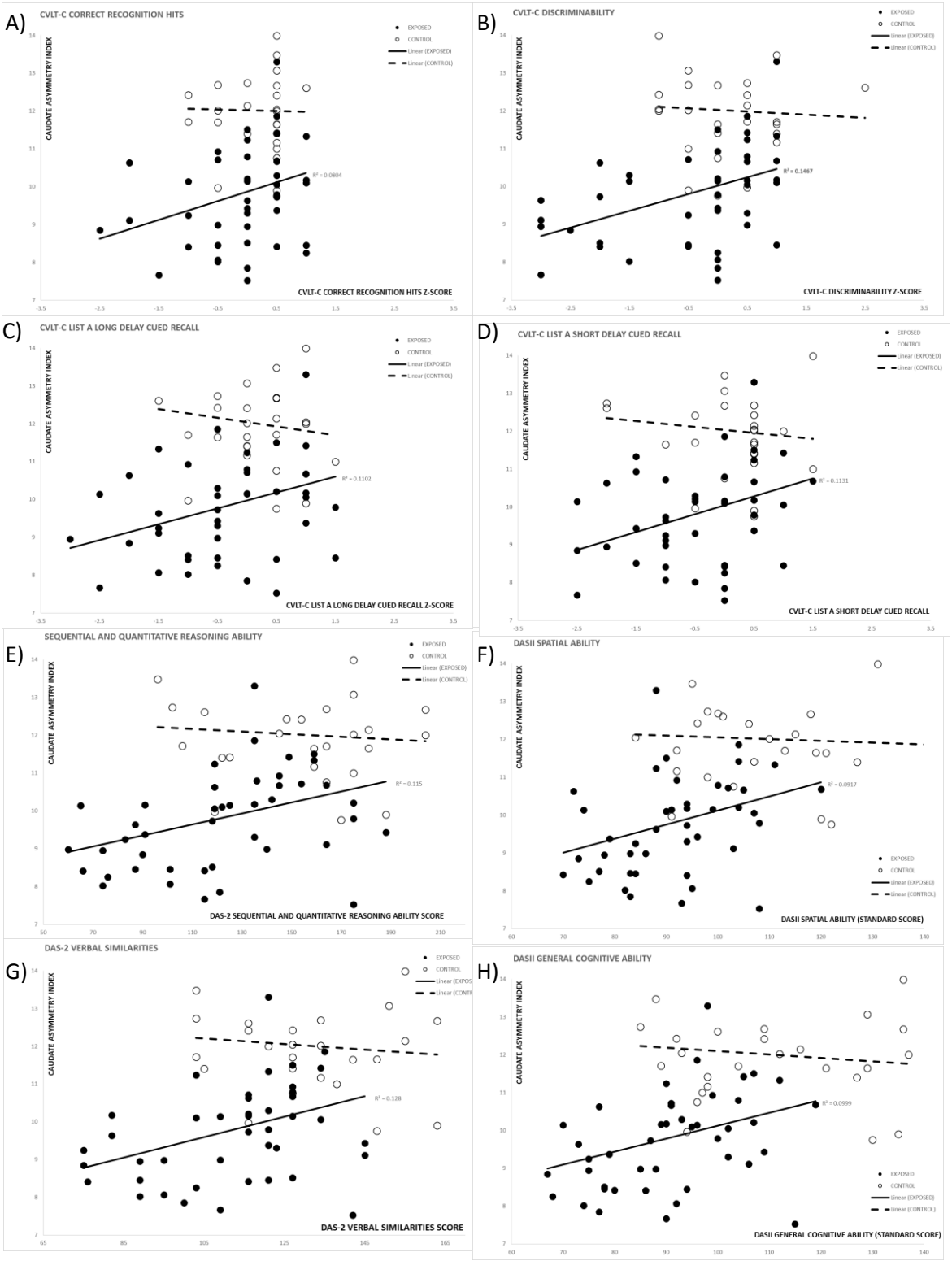


Figure 5.16 Caudate nucleus asymmetry index for exposed (FAS+HE; n=44) and control (n=26) groups plotted against age-normalised neurocognitive scores. Showing positive correlations in the exposed groups for CVLT-C (A) correct recognition hits (r^2=0.08, p=0.065), (B) discriminability (r^2=0.13, p<0.02), (C) Long delay cued recall (r^2=0.11, p<0.03), (D) short delay cued recall (r^2=0.13, p<0.03). Similarly, DAS-2 standardised scores show significant correlations for (E) sequential and quantitative reasoning ability score (r^2=0.11, p<0.03); (F) spatial ability (r^2=0.09, p<0.05), (G) verbal similarities ability score (r^2=0.13, p<0.02); and, (H) general cognitive ability (GCA) (r^2=0.1, p<0.04).

5.5 Discussion

The deleterious and disproportionate effect prenatal alcohol exposure is thought to have on corpus callosum (CC) development (McGee, 2005) gives a valid rationale for its selection as a component to evaluate combined face-brain techniques. The CC is the largest myelinated midline structure in the brain primarily responsible for interhe#mispheric communication. Early autopsy studies focusing on the effects of maternal alcohol consumption at dangerous levels to cause cell migration disruption, full agenesis or thinning of the corpus callosum, abnormalities in the brain stem and cerebellum (Jones and Smith, 1973; Clarren and Smith, 1978b). More recent studies utilising MRI imaging have documented reduced callosal volume (Yang et al. 2012; Sowell et al 2001), displacement of the posterior corpus callosum regions (Sowell et al 2001), morphological anomalies (Bookstein *et al.*, 2007) and localised reduction in callosal thickness (Yang et al. 2012). Additionally, the presence of partial or full agenesis of the CC in individuals with PAE is thought to be in the order of 6.8% (Riley *et al.*, 1995b). The resulting disruption in CC development has been associated with a range of neurocognitive and neurobehavioral deficiencies in learning, verbal memory, executive function, social cognition and attention (Fryer *et al.*, 2012).

A previous study of face-brain relationships in a murine model of prenatal alcohol exposure (Lipinski *et al.*, 2012) used a DSM approach to analyse the shape of the face and brain components simultaneously. The face and brain components were segmented simultaneously from MRI volumes and thus could be processed in a single DSM model. In theory, this is possible with MRI volumes of the human head but the compression pads used to restrict head movement distort the soft tissues of the face. Therefore, in this study we used an existing dataset of high-resolution facial images acquired with 3D stereophotogrammetry and segmented structural MRI images from 3T scanners of the same individuals and employed separate face and brain surface representations to produce a series of combined models, each with a single set of PCA outputs. The resulting models represent the morphometric covariation of each brain component and the facial region linked statistically by a series of amalgamated principle component (PC) values. Additionally, we focus not only on the whole face but specific areas, which in FASDs are shown to have greater diagnostic significance. Previous DSM studies (Hammond *et al.*, 2005; Chinthapalli *et al.*, 2012; Hammond, Hannes, *et al.*, 2012; Suttie *et al.*, 2013, 2017) have shown accurate discriminating power of using only localised regions of the face.

In this analysis, midline facial regions (profile, philtrum, nose and upper-lip vermilions) and non-midline regions (face, eyes, mandible and malar) combined with the corpus callosum, and caudate nucleus in some instances, is more concordant with clinical classification for a FAS-control dichotomy. Midline regions combined with a midsagittal representation of the corpus callosum performed better the non-midline facial regions, improving FAS-control classification testing when compared to single structure models. Combining the caudate nucleus with the face did not improve classification accuracy, but when dividing the structure into its respective left-right halves, we determined noteworthy improvements using the left caudate combined with all regions except the nose. In all combined testing, we demonstrated that both brain structures improve control-FAS classification accuracy when combined with the lip-vermillion. This could perhaps be due to lack of discriminatory information provided by the lip-vermillion phenotype.

Exposed populations show a significant reduction in bilateral caudate nucleus volume as a proportional decrease relative to intracranial volume (Archibald *et al.*, 2001; Astley *et al.*, 2009;

Joseph *et al.*, 2014). It is evident from our PC1 representations and calculated volumes of the caudate nucleus that growth is diminished in exposed populations. However, our models do not take into account intracranial volume. As reported in the literature, the corpus callosum does not show consistent volume change with age (Pfefferbaum *et al.*, 2013). Likewise, in our analysis, growth trajectories from PC values did not correspond with the associated face-age trajectory (Suttie et al. 2013).

Given that the combined classification testing for left and right caudate differed so much it is fair to hypothesise left-right asymmetry differences between FAS and control groups. Using a methodology previously utilised in DSM studies of facial asymmetry in Autism Spectrum Disorders (Hammond *et al.*, 2008) and the effects of different cancer treatment strategies (Schoot *et al.*, 2017), we calculated measures of asymmetry for the caudate nucleus which on average shows control caudate nucleus to be more asymmetric compared to the FAS group. This is not the case for the HEs, who show no significant asymmetry index difference from controls. A displacement heat map depicting axial asymmetry along a depth axis shows a left anterior dominance in controls and similarly in the FAS group but to a much lesser extent. In the HE group, the direction of asymmetry is inconclusive, indicating heterogeneous asymmetry across the group.

In a previous study, Sowell et al. 2002 reported individuals with PAE to have reduced asymmetry of the cortical surface and grey matter density compared to controls, indicating that the development of typical brain asymmetry can be lost in exposed populations. As craniofacial development is intrinsically linked to the brain, it seems natural for asymmetry of one to be reflected in the other. However, contrary to our findings and those of Sowell et al., a study of facial asymmetry (Klingenberg *et al.*, 2010) observed significant differences in axial asymmetry of individuals with a FAS diagnosis compared to controls, in both a Finnish and a South African population. In this landmark based study, average axial asymmetry was identified as a midline landmark shift to the right and of the eyes to the left. In our study of the caudate nucleus, it is perhaps the deep location of this structure within the brain that disengages its relationship with the midline of the face. There is a higher likelihood that asymmetry of the brain will influence the face when regions of focus are directly adjacent (Hammond et al. 2008).

When testing the caudate asymmetry relationship to neurocognitive measures, we found the exposed population to express significant correlations on specific CVLT-C and DAS-2 outcomes. Variables responsible for assessing verbal learning, delayed recall, general cognitive and spatial ability showed a positive, and significant, correlation with caudate asymmetry index which was not apparent in the control group. Similar neurocognitive measures have been used in previous studies of the caudate. Fryer et al. 2012 investigated relationships of brain structure volumes and neurocognitive outcomes in 21 individuals with histories of PAE. Among the components analysed in this study was the caudate nucleus, which emerged as the most consistent predictor of cognitive performance in the PAE group. The authors utilised CVLT-C testing for verbal learning and recall scores finding strong correlations with the bilateral volume of the caudate nuclei.

Combining the whole caudate did not vastly improve classification accuracy, but when dividing the structure into its respective left-right halves, we see noteworthy improvements using the left caudate combined with all regions except the nose. Given the combined classification testing for left and right caudate differed to such a degree it is fair to assume the presence of right-left

asymmetry differences between FAS and control groups. Using a methodology previously utilised in DSM studies of facial asymmetry in Autism Spectrum Disorders (Hammond *et al.*, 2008), we calculated measures of asymmetry expressed in the caudate nucleus. The calculated index characterises asymmetry of the left-right brain morphometry that on average shows controls to possess a significantly higher degree of asymmetry when compared to the FAS group. This is not the case for the HEs, who show no significant differences for asymmetry index compared to controls.

Previous studies of the caudate nuclei, with mention of left-right differences predominantly focus on volumetric asymmetry (Hynd *et al.*, 1993; Archibald *et al.*, 2001; Astley *et al.*, 2009; Willford *et al.*, 2010; Nardelli *et al.*, 2011; Fryer *et al.*, 2012; O'Dwyer *et al.*, 2016). Willford et al. produced a study focusing on asymmetry of the caudate nucleus in offspring of mothers who drank alcohol in moderate quantities during pregnancy. A population of young adults (n=45; 18-22yrs) were recruited for this MRI study and separated into 3 groups based on their trimester dependent prenatal alcohol exposure; controls (no alcohol exposure, n=20), 1T (first trimester only exposure, n=15) and 3T (exposure throughout all three trimesters, n=10). The authors calculated the magnitude of asymmetry by subtracting the left caudate volume with the right and dividing by the total ([L-R]/[L+R]). The resulting analysis found dose-dependent effects with the 3T groups showing a statistically significant leftward bias in caudate asymmetry compared to 0T, with the 1T group being an intermediate between the 0T and 3T groups but statistically non-significant compared to either. Additionally, this study controlled for tobacco and marijuana use concluding that neither of these significantly affect caudate volume or asymmetry. Although this study was through in controlling for influencing factors, results were limited by small sample size reducing the statistical power.

That only the FAS group demonstrated a significant, and more consistently reduced left-right caudate nucleus asymmetry compared to controls is most likely due to the fact, that to create the craniofacial features of FAS, alcohol exposure must occur in the 3^{rd}-4^{th} weeks of human development. At this stage, the tissues that will make up the brain and face are still developing as a single structure, which is why craniofacial dysmorphology frequently accompanies striatal and other rostral midline brain structure abnormalities (Godin *et al.*, 2010, 2011; Donald *et al.*, 2015). The less consistent effects in the HE group suggests alcohol exposure during later stages of development, demonstrating that while the severe craniofacial effects in FAS children largely result from prenatal alcohol exposure in a narrow time window, while the caudate nucleus (and many other regions of the brain) are susceptible over a much wider period. The fact that ethanol exposure seems to have a more significant effect on the left caudate nucleus has also been observed in a mouse model of early gestational alcohol exposure, where both sides of the caudate nucleus was affected, but there appears to be a slight leftward bias, possibly due to the differential left-right sonic hedgehog (Shh) concentration gradients present in early embryos (Fish *et al.*, 2016).

As with the previous chapters, this study accentuates HE individuals. We seek to understand if the combination of face-brain abnormalities provides a means to identify associated biomarkers exploited to improve diagnostic classification for this commonly underdiagnosed subset. Our main objective focused on determining if combining face and brain features, using a single joint model, could improve analysis and diagnostic outcome for individuals across the FASD spectrum.

5.6 Conclusions

Determining correlations between localised facial regions and structural brain anomalies is important as it could ultimately provide insight for clinicians to determine the extent of neuro-structural damage simply by examining the face. It is important to emphasise that MRI image analysis is highly unlikely to ever be part of the routine diagnosis of FASDs. However, its use in face-brain analysis might help identify subtle facial dysmorphism associated with neurological pathologies in FASD individuals, particularly the HE population where facial recognition is challenging.

Further investigation is required into the neurocognitive measures and their relationship with face brain morphology. Working in collaboration with CIFASD we hope to obtain detailed neurocognitive assessment scores that accompany this dataset. We also wish to investigate the capability of this new technique with alternative regions of the brain. A recent study using a form of morphometric shape analysis has found specific alcohol-related deficits in the hippocampi and caudate nuclei regions. An initial investigation into the MRI segmented data available to us reveals that we can extract caudate nucleus 3D surface geometry. However, the DSM process requires anatomically accurate landmarking of structures prior to processing to induce a dense surface point correspondence. The landmarking process on such structures is laborious, time-consuming and prone to error. Working in collaboration with the authors of this investigation and an expert imaging group at UCL, we hope to automate this process enabling single or combined DSM techniques to be applied to any neurological structure in combination with the face.

6 CLINICAL TRANSLATION

In this chapter, we outline a clinical protocol for identifying FASD associated facial features based on the findings from the preceding chapters. A fundamental aim is to develop a facial screening tool to introduce into the clinical workflow. Development and testing were carried out in collaboration with a national FASD clinic, with clinical expertise strongly contributing to improved iterations of the software. Within this chapter, we disassemble each of the techniques included and introduce the resulting software package 'FaceScreen'

6.1 Introduction

The initial chapters of this thesis have focused on identifying and accurately quantifying facial features in FASDs. Using both existing and novel additions to the DSM toolkit we have achieved accurate control-FAS discrimination, identified face-brain relationships and recognised face-brain and neurocognitive correlates that exists in PAE populations. DSM analysis has repeatability proved to deliver a powerful set of analytical tools for human facial (Hammond *et al.*, 2005; Chinthapalli *et al.*, 2012; Hammond and Suttie, 2012; Hammond, Hannes, *et al.*, 2012; Suttie *et al.*, 2013) and murine face-brain analysis (Lipinski *et al.*, 2012). So far, several key enhancements have been made throughout this project for the FASD-specific application of DSM analysis tools, but have been confined to a research environment. While a DSM approach to the identification of FASD associated facial features has proven to be effective, the research orientated software environment has limitations regarding usability and functionality, rendering it unsuitable as a clinical tool without development and modification.

The ongoing challenge remains to be unbiased, objective evaluation of facial features which are currently assessed by a combination of measurement and judgement based on the skill and experience of the clinician. In this chapter, we investigate how facial evaluation contributes to clinical evaluation in a FASD clinic, and how the screening of 3D face shape might be beneficial. The National Clinic for Foetal Alcohol Spectrum Disorders (http://www.fasdclinic.com) is an NHS-funded clinic specialising in the assessment and treatment of children and adults with FASD. We have worked throughout this phase of the project at this clinic, in close collaboration with

the leading UK FASD clinician Dr Raja Mukherjee, to develop and trial 3D facial screening as part of the clinical protocol for the recognition of facial features in individuals with suspected FASD. Attendance at this national clinic offered insight into the requirements and demands of a front-line environment. With continuing clinical interaction and contributions from Dr Raja Mukherjee, we aim to distribute these software tools for use in FASD clinics worldwide.

In previous chapters, we identified a series of facial regions that alone provide accurate control-FAS discrimination rates. In particular, in a Caucasian cohort the malar, nose, eyes, mandible and philtrum all performed well. In the diagnostic guidelines, reduced PFL, thin vermillion border and a smooth philtrum are referred to as cardinal FAS features. Features we identified as useful (flattening of the malar, retraction of the mandible (retrognathia), flattening of the nasal bridge, and a short button like nose) are typically categorised as minor anomalies (Hoyme et al. 2016). The Hoyme and other diagnostics guidelines only take into consideration the 3 cardinal facial features when assigning a diagnosis, omitting the remaining minor anomalies from any clinical categorisation (Coles *et al.*, 2016). This could, in part, be due to the difficulty in accurately identifying and quantifying such features. For example, using image process techniques, angular or linear landmark based measurements could be utilised to assess the prominence of the nose and mandible, but manually taking these measurements would be infeasible. DSM based analysis has allowed us to focus predominantly on the features not typically used for diagnosis and determine their significance within the heavily exposed individuals without the required criteria for a FAS or PFAS diagnosis.

6.2 Clinical Protocol

Patients presented at Dr Mukhejee's clinic arrive on a referral basis and are considered from the age of 6 upwards. An individual under scrutiny would typically have been brought to the attention of care professionals, or by guardians or parents, at early stages of education when behavioural and neurodevelopmental deficiencies become apparent. In many cases, educational support and behavioural intervention are already likely to be in place. A subsequent assessment by clinical geneticists is a prerequisite for referral, where genetic testing has ruled out the possibility of differential diagnoses. Typically, prenatal alcohol exposure information will either have been confirmed or strongly suspected, however, there are some instances where exposure remains unknown. In these cases, a diagnostic challenge arises and it is the job of the clinician to rule out any differential diagnosis. Furthermore, if the patient has a presentation in keeping with FASD criteria, but without the history of prenatal alcohol exposure, it would not be possible to confirm a FASD diagnosis. If confirmed exposure information is available in the future, it is possible to revisit this conclusion.

Dr Mukhejee's clinic sessions are split across two days, the first of which requires the attendance of the patient and an accompanying parent, guardian or main caregiver. The initial session involves a ~ 90-minute assessment determining patient history and facial analysis using the FAS Facial Photographic Analysis Software in combination with the 4-Digit Diagnostic Code (Astley and Clarren, 2001) (see section 2.2.2). The latter part of the day involves the patient completing a series of neuropsychological and communication assessments, typically lasting 3-4 hours in which a battery of tests builds a neurocognitive and behavioural profile. The parent or guardian accompanying the patient is given a series of assessments to be completed by the person closest to the patient detailing further background: written confirmation and any known detail of alcohol exposure and criminal involvement; teacher assessments; and, behavioural

characteristics. For children, the day 2 assessment requires only an informant, often a parent or guardian, and consists of a neurodevelopmental assessment and confirmation of the information gathered from the day 1 sessions. Figure 6.1 details the patient evaluation and referral process that patients will undergo at The National Clinic for Foetal Alcohol Spectrum Disorders.

Figure 6.1 The National Clinic for Foetal Alcohol Spectrum Disorders patient referral and evaluation process (*source: http://www.FASdclinic.com/referrals*)

6.3 Objectives

We have shown in previous chapters the ability to distinguish control and individuals with a FAS/PFAS diagnosis accurately from one another. However, as previously stated – the vast majority of individuals with PAE do not express facial features required for a FAS/PFAS diagnosis. At the UK FASD clinic, attendees predominantly fit into this category. Our primary objective is to provide a clinical tool for the recognition of facial anomalies that can be attributed to PAE. The clinical criteria for FASD diagnoses (Hoyme *et al.*, 2016) identify several minor facial anomalies associated with PAE that do not form part of the FAS or PFAS diagnoses. Features such as hypoplastic midface, decreased interpupillary and inner canthi distance (≤25th percentile), long philtrum (≤90th percentile), anteverted nares and a flat nasal bridge are all recognised but do not form any part of diagnostic criteria. To cover the spectrum of associated facial anomalies, we can break down our objectives into two sections; identification of the three cardinal features (smooth philtrum, shortened PFL and thin vermillion) and the identification of minor anomalies listed above. Thus, the development of a clinically useful tool, embodying results of previous analysis and providing an objective assessment of the presence of facial anomalies must be able to provide the following:

a) Guided Landmark Placement

DSMs, anthropometric measurements and directional curvature calculation all require the placement of a series of anthropometric landmarks. This will be a requirement of the clinician, so this implementation must provide a guided method for landmark placement to maintain accuracy and consistency.

b) Objective identification of philtrum smoothness

Current methods using Likert scale are subjective, philtrum assessment using an image-guided process is required to be accurate, reliable and quantitative.

c) Accurate Palpebral Fissure Length Measurement

As a cardinal feature of FAS and PFAS, we require accurate measurements of PFL to be calculated and normalised against an age-sex matched control dataset.

d) Control-FAS Classification

It is a requirement of this tool to determine if an individual has a facial presentation in fitting with FAS. To achieve this, we will need to test an individual against an optimised set of training data containing previously recruited control and FAS subsets.

e) Identification of Minor Abnormalities

Although not forming part of the diagnostic criteria, it is beneficial in an evaluation to record the presence or absence of these features in order to identify all minor anomalies associated with PAE (Hoyme *et al.*, 2016). Landmarks can quantify several features mentioned above, or be recognised using DSM models depicting separate regions of the face. Regional classification models have proven to accurately discriminate between controls and individuals with FAS/PFAS. So the inclusion of regional models such as the malar region, philtrum, nose and eye models can provide some indication of these localised facial anomalies.

f) Clinical Report Generation

For this tool to be a useful part of the clinical workflow, all the prior mentioned objectives should corroborate to specify the likelihood that the facial presentation of an individual is associated with PAE. This needs to be summarised into an automatically generated report providing an objective facial assessment.

6.4 Methods

6.4.1 Datasets

In this chapter, we utilise previously analysed datasets from chapters 3, 4 and those used in previously published studies (Suttie *et al.*, 2013, 2017). To recap; the South African, Cape Coloured cohort utilised (Chapter 3) included 192 participants containing the exposed groups FAS (n=22), PFAS and HE (n=75). The CIFASD recruited cohort (Chapter 4) consisted of 249 Caucasian participants recruited from the United States (n=207) and several countries in Europe (n=42). We excluded participants who self-reported as Hispanic to maintain uniformity. Diagnostic categorisation in the Caucasian cohort did not use the PFAS label, but rather the exposed groups FAS (n=33) and HE (n=73). Using an alternative hand-held imaging system (Vectra H1) we test only PFL measurements using 3D images of 36 CIFASD, and 10 Caucasian patients from the UK FASD Clinic. Patients from the UK FASD cohort were used to generate clinical reports for the purpose of clinical validation of 3D facial screening.

6.4.2 3D Image Acquisition

In this study, we use the Vectra H1 3D Imaging System from Canfield Scientific (http://www.canfieldsci.com) for capturing 3D photographs of the face (Figure 6.2). This is a commercially available hand-held stereophotogrammetry device, capable of capturing high-resolution 3D images with a geometric resolution of 0.95 mm (triangle edge length). This system is built from a high-resolution Canon SLR body with a bespoke mirrored stereo-photographic lens, providing a split image with dual perspective for depth recognition. Previous 3D studies conducted for DSM analysis have utilised static, tripod-mounted camera systems which due to cost, physical dimensions and portability are not ideal for clinical deployment. The cost of H1 devices is typically ~£8,000 per unit, making it an economical choice compared to the static units on the market at around £30,000.

Figure 6.2 Vectra H1 3D Imaging System from Canfield. Constructed from a relatively inexpensive Canon EOS Rebel T3i SLR camera body with a specialised stereophotogrammetric lens. A soft light flash and green-light guided range lights ensure a well-lit image taken at the correct distance for accurate 3D reconstruction.

Three separate images are captured and stitched together to form a uniform mesh of 3D points (Figure 6.3). The subject is required to be still with a neutral facial expression which must remain fixed for each of the three captures. Software provided by the manufacturer guides the user for the positioning of each image; lower right, central and lower left (Figure 6.3; A, B, C). Once complete the automated stitching process takes 1-3 minutes depending on image size and processing power available. The acquisition process requires a degree of compliance from the subject and can obtain capture errors in the more dynamic areas of the face such as the lips and eyes if changes in expression occur between the three images.

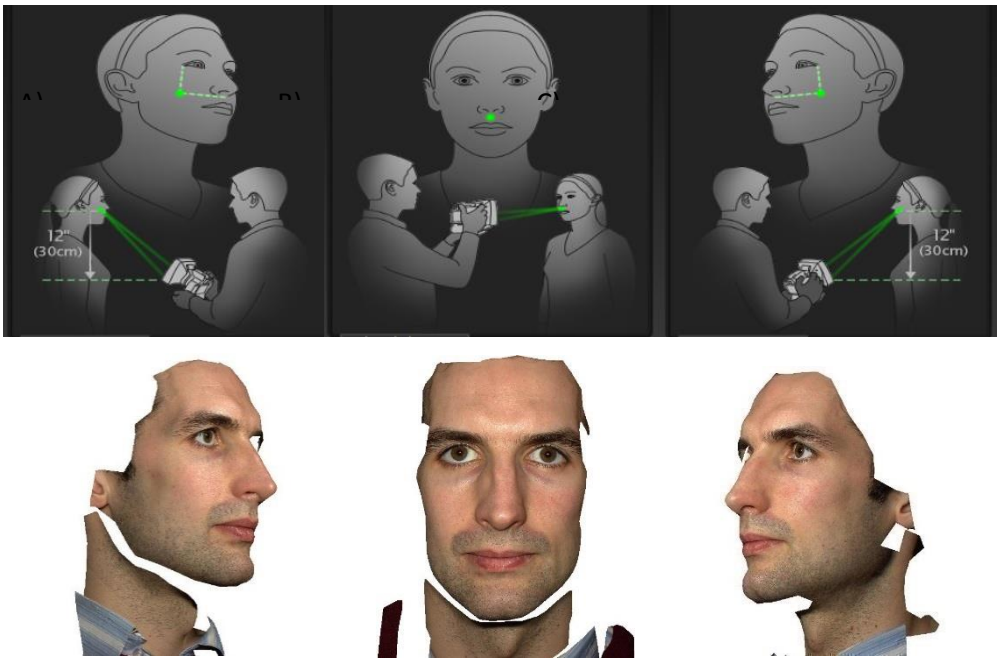


Figure 6.3. The three stages of image capture from lower right (A), central (B) and lower left (C) positions using the H1 Vectra 3D Imaging System. Image courtesy of Canfield Scientific (www.canfieldscientific.com)

6.4.3 Anthropometric Landmark Placement

A series of reliable anatomical landmarks is required before the application of the techniques detailed in this section. Prior analysis has involved manually placing landmarks by experienced users. A clinician will not have the comparable expertise, so it is important to test the reliability of placement, and ensure there is a balance between the number of landmarks required and the performance of methods employed. Typically, in previous analysis (Chapters 3-5), 24 manually placed points were required to support an accurate synthesis of a face surface in a DSM model. An experienced user can accurately place 24 landmarks in a matter of minutes, but a clinician during a patient session will not have the time or experience to precisely place so many points. We propose and test a reduced series of 13 landmarks (Figure 6.4) that adequately provide surface point correspondence and required anthropometric measures without compromising the model surface synthesis and discrimination ability. These include: 5 midline points for midline defects; inner and outer canthi of the eyes for PFL; lip centre and cupid points for philtrum smoothness; and, lower points of the ears for lateral positioning.

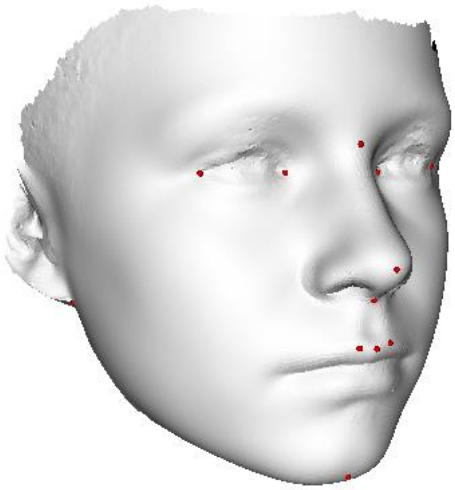


Figure 6.4 Placement of a reduced set of 13 landmarks used for point correspondence. Left and right: exocanthion, endocanthion, otobasion inferius and crista philtrum. Mid-line points: Nasion, gnathion, subnasale, pronasale and labiale superius.

PFL measurement is calculated as the mean of the left and right Euclidean distances between their respective exocanti and endocanthi points. To calculate percentiles and Z-score PFL measurements, we utilise normative values from a database containing 2454 male and female individuals of European Caucasian ancestry aged 3-40. PFL values were obtained from manually positioned eye points on 3D images, obtained from FaceBase (www.facebase.org) generated by project U01DE020076. The FaceBase consortium and the FaceBase Coordinating Hub (1U01DE024449-01) are funded by the National Institute of Dental and Craniofacial Research.

For the assessment of the philtrum, curvature methods described in the proceeding section require axes defined by a series of 2 lateral and 2 horizontal landmarks. In this case, it makes sense to utilise left and right eye points (exocanthi points in this case) for the horizontal axis, and subnasale and upper lip centre points for the vertical axis. This vertical axis is preferable to the use of nasion and gnathion points as it is representative of the angle of the philtrum.

6.4.4 Philtrum Assessment

Throughout this thesis, we have shown the potential to identify localised indicators of facial dysmorphology using individual 3D images normalised against age-matched controls. Normalised heat maps derived from this method (face signatures) provide key indicators of surface displacement relative to a control population. However, it is the case that some individuals severely afflicted with growth deficits will have facial signatures principally showing size differences, diluting any other minor facial anomalies such as philtral smoothness (Figure 6.5B, col. 2). In chapter 4, we looked at philtrum assessment using normalised curvature of the philtrum (Figure 4.4) and concluded that curvature heat maps are a better indicator of philtral smoothness than displacement signatures. For clinical identification, mean overall curvature of the face (Figure 6.5A and B, col. 3) and the previously described directional “Curl” and “Groove” curvatures (chapter. 4, Figure 4.3; Suttie et al. 2017) can be instantaneously calculated on the

whole face with the addition of 4 landmarks that can be used to define axes. This calculation will be included in the software and immediately accessible to the user.

6.4.5 Classification Testing to Identify PAE Associated Facial Features

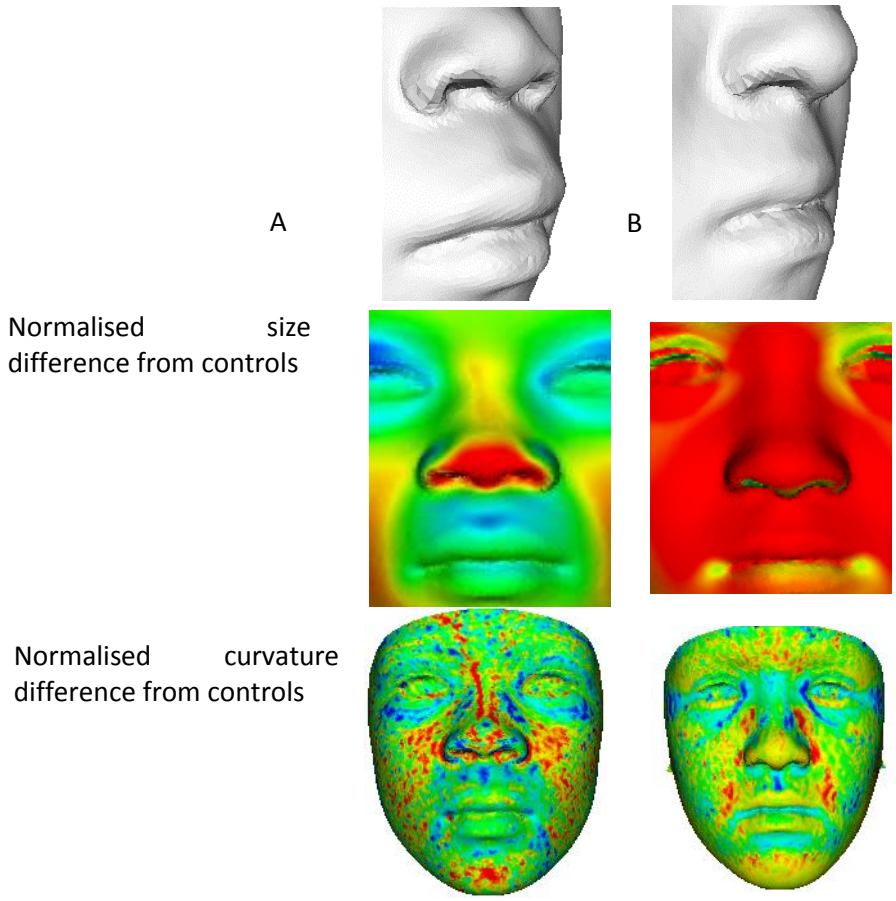


Figure 6.5. Two individuals with a FAS diagnosis, both with smooth philtrums, when visualising facial signatures (column 2) at ±2SD. The signature of face A shows localised blue colouring on the philtrum indicating an expansion in this region, whereas colouring on face B's philtral smoothness is diluted by red (reduction) due to a highly significant reduction in facial growth. Column 3 shows normalised mean curvatures (±1.5SD) for each face, in both individuals we observe a region of blue colouring in the philtrum indicating a convexity of the philtrum groove.

In the previous chapters, we have tested agreement of control-FAS discrimination based only on shape using a series of classification algorithms. In each case for the Caucasian cohort, the closest mean (CM) classifier performed equally well compared the alternative algorithms; support vector machines (SVM) and linear discriminant analysis (LDA). For this reason, and one of efficiency (LDA and SVM algorithms can be computationally demanding), we choose the CM classifier for this implementation. Using the same principle as previous analysis, we divide our training set across 20 random 90% splits. An individual with landmarks concordant with those of the classifier can be tested to determine their position relative to the control and FAS means for each spilt. The resulting decision is calculated as the mean position across 20 splits for which we also calculate the standard error. To identify cardinal FAS features, we apply this to the whole face, profile, philtrum and upper lip vermillion regions given the accuracy previously seen in the analysis of the Caucasian cohort. Minor anomalies can also be detected by utilising mandible, malar, midline profile and nose models.

Previous analysis utilised a series of 24 manually annotated landmarks for classification testing. This would be an expectation of the clinician, so reducing these is key to avoiding usability issues. We repeat previous control-FAS testing using a reduced landmark set of 13 landmarks (Figure 6.4) to determine if classification accuracy can be maintained. To test classification model accuracy, we discriminate between sets of unseen randomly sampled test examples across 20 random 90% to 10% training-unseen test pairs using DSM models built with a reduced set of landmarks. As with previous analysis, we calculate discrimination rates between control and FAS subsets in terms of the mean area under the receiver operating curves (ROC) of the 20 trials, representing the probability of correctly classifying a pair of randomly selected faces with each of the pairs from a different subgroup.

Classification models to be used in the final implementation are built using 100% (n-1) of PCs, which is then optimised by reducing the number of PCs to the point where classification accuracy is at its highest. Typically, in previous DSM studies, models are built using 99% of PCs to cover face shape variation (Hammond *et al.*, 2005). However, for classification models of conditions with distinct features such as in FAS, there will be an optimal number of PCs that will cover shape variation required to delineate associated features. By calculating classification accuracy at every accumulation of PCs, we can determine where to set a cut off for each training model.

Figure 6.6 shows the result of testing classification accuracy at each PC for control-FAS discrimination testing of the nose. In this example, the optimal number of PCs is 27, representing 98.2% of shape variation. We can deduce that individual surface synthesis at this point is sufficient to accurately represent discriminating morphology without introducing noise.

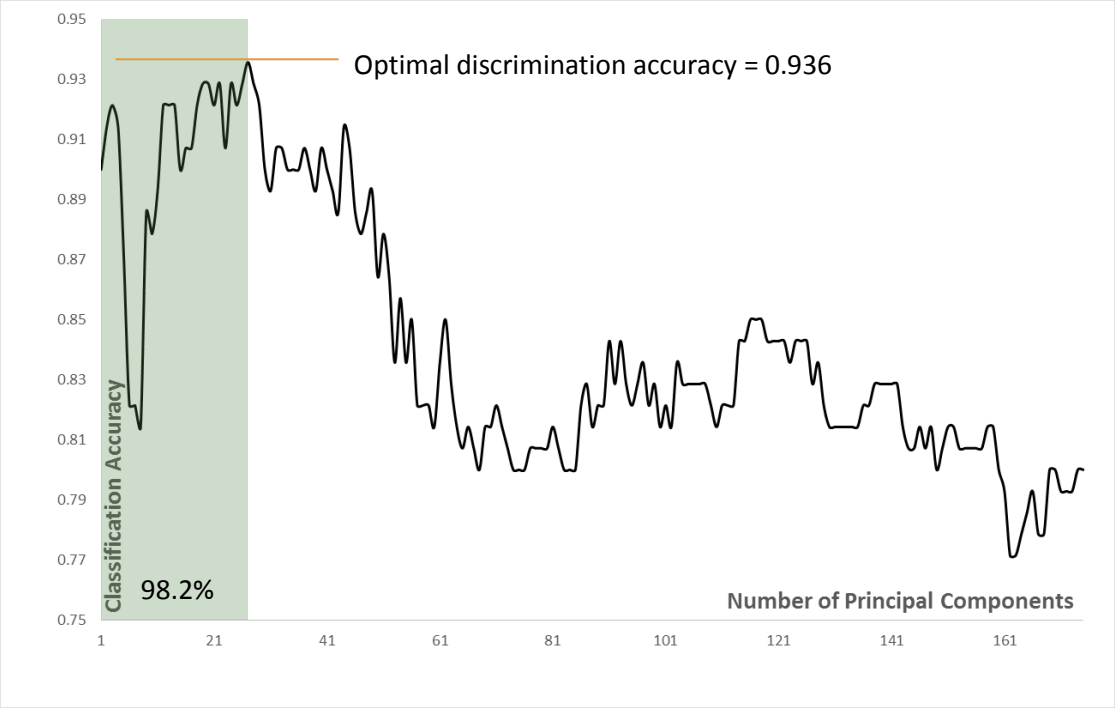


Figure 6.6 Control-FAS classification accuracy at each PC for a nose model. We discriminate between sets of unseen randomly sampled test examples across 20 random 90% to 10% training-unseen test pairs using DSM models of the nose. As with previous analysis we calculate discrimination rates between control and FAS subsets in terms of the mean area under the receiver operating curves (ROC) of the 20 trials,

classification accuracy represents the probability of correctly classifying a pair of randomly selected faces with each of the pairs from a different subgroup.

6.4.6 Identifying Facial Anomalies

Using the clinical software, we aim to objectively identify as many of the facial features associated with PAE as possible. We achieve this with a multifaceted approach, utilising the most suitable tested techniques and applying them to an individual subject. All information for each individual is required to be compiled into a report, to indicate the likelihood that a patients' presentation is associated with PAE. Minor anomalies can assist in this overall assessment of the face as mentioned by Hoyme and colleagues (Hoyme et al. 2016), and go beyond the typical cardinal features. The Hoyme guidelines provide a dysmorphology scoring system based on the assessment of minor anomalies in 370 children with FAS. Each facial anomaly is ranked 1-3, where the cardinal features were assigned a score of 3. Features that appeared in ≥100 children were assigned a score 2 and those observed in <100 were assigned a score of 1. Table 6.1 summarises the facial anomalies as listed in the diagnostic guidelines, the current method used for diagnosis, their assigned scores and our proposed method for their recognition. From the analysis in earlier chapters, we identified certain non-cardinal features using region-specific models as being particularly useful for identifying individuals with PAE. These are inherently linked to the facial anomalies listed in Table 6.1. For example, closest mean control-FAS classification accuracies achieved in our Caucasian cohorts for hypoplastic midface (malar classification model) 0.90; flat nasal bridge and anteverted nares (coverage by nose classification model) 0.94; and prognathism (mandible classification model) 0.86.

Anomaly	Current Method of Identification	Proposed Method	Hoyme et al. 2016 Score
Smooth philtrum	Lip/Philtrum Guide (Astley and Clarren, 2000; Hoyme *et al.*, 2016)	DSM Classifier Facial Signature Normalised Curvature	3
Thin lip vermillion	Lip/Philtrum Guide (Astley and Clarren, 2000; Hoyme *et al.*, 2016) Lip circularity from 2D image (Astley and Clarren, 2000)	DSM Classifier Curvature analysis	3
Reduced PFL (≤10th percentile)	Manual measurement (Hoyme et al. 2016) Measured from 2D images (Astley and Clarren, 2000)	DSM Classifier 3D Landmark Based Measurement	3
Hypoplastic midface	Likert chart scoring 1-3 (Astley and Clarren, 1996)	DSM Classifier Facial Signature	2
Philtrum length (≥90th percentile)	Manual measurement	3D Landmark Based Measurement	2
Decreased ICD/ICD (≤25th percentile)	Manual measurement	3D Landmark Based Measurement	2
Anteverted nares	Clinically identified as present/not present	DSM Classifier	2
Flat nasal bridge	Clinically identified as present/not present	DSM Classifier	2
Prognathism	Clinically identified as present/not present	DSM Classifier	1

Table 6.1 Facial features as listed in the updated guidelines for diagnosing foetal alcohol spectrum disorders, the current methods used for identifying them, the methods proposed to recognise them using our 3D facial analysis, and the score given by the diagnostic guidelines (Hoyme et al. 2016).

6.5 Results

6.5.1 3D Image-Based Palpebral Fissure Length Measurement

During assessment in FASD clinics, PFL is recorded either with a ruler by eye or using the 2D photography method (Astley and Clarren, 1996). PFL estimates taken in our background test datasets from CIFASD (Caucasian) and the University of Cape Town (Cape Coloured) used ruler measurements during clinic assessment. Using manually placed eye landmarks on 3D images to calculate PFL resulted in significantly greater values compared to using a ruler for: Caucasian (26.48±2.0 vs 25.34±2.06; $p<0.00000002$; Figure 6.7A); and Cape Coloured (26.4±1.7 vs 25.3±1.9mm; $p<0.00000001$; Figure 6.7B) cohorts. To calculate percentile rank using ruler measurement, the mean of the left and right PFL lengths are compared to sets of normative PFL charts obtained from similarly measured controls. The same charts cannot be utilised for 3D measurements given the disparity between 3D and ruler PFL values. Instead, we use normative data from eye points manually placed on 3D images obtained from FaceBase (www.facebase.org). Testing the PFL from the FaceBase dataset (n=2454) against the PFL obtained from the H1 camera on a small Caucasian subset (n=36; Figure 6.7C) and from static tripod mounted systems (n=249; Figure 6.7D) we find no significant differences.

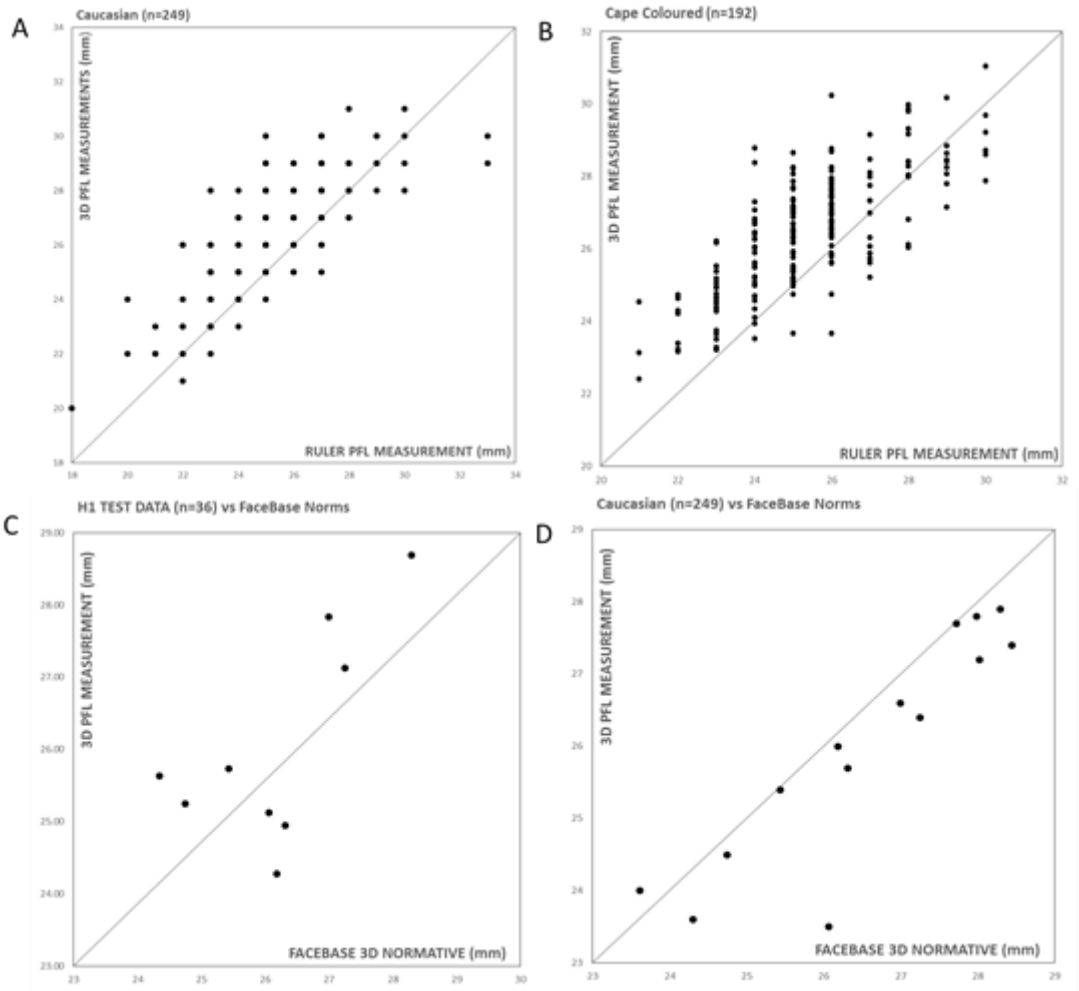


Figure 6.7 PFL measures from 3D camera systems compared to measures obtained using a ruler in A) Caucasian cohort (n=249); B) Cape Coloured (n=192). C) Demonstrates the use of the H1 camera using a small Caucasian test (n=36) and D) Caucasian CIFASD set using static 3D cameras showing uniformity with FaceBase normative PFL values.

6.5.2 Control-FAS Discrimination and Number of Landmarks

The task of manually placing 24 landmarks accurately on each subject for DSM model construction is an unreasonable expectation for a clinician. 24 landmarks produce an accurate DSM representation of the face, but there may be facial regions where accurate surface synthesis is redundant or can be sufficiently obtained with fewer landmarks. We reproduced the control-FAS discrimination testing from Chapters 3-5, using a reduced set of 13 landmarks, for the full face and each of the 7 facial regions (nose, malar, midline profile, mandible, eyes, philtrum and upper-lip vermillion). We optimise classification models by reducing the number of PCs to the point where closest mean classification accuracy is at a peak. Using models built from surface correspondence calculated using a reduced set of 13 landmarks does not appear to affect classification accuracy (Table 6.2), surface synthesis or normalised heat maps (Figure 6.8).

	Closest Mean	
	24 landmarks	**13 landmarks**
Face	0.95	0.94
Profile	0.83	0.84
Eyes	0.92	0.87
Malar	0.90	0.90
Mandible	0.86	0.86
Nose	0.94	0.94
Lip vermilion	0.70	0.70
Philtrum	0.83	0.81

Table 6.2 Discrimination rates for 24 and 13 landmark classification models. Proving that little accuracy is lost when reducing the landmark set.

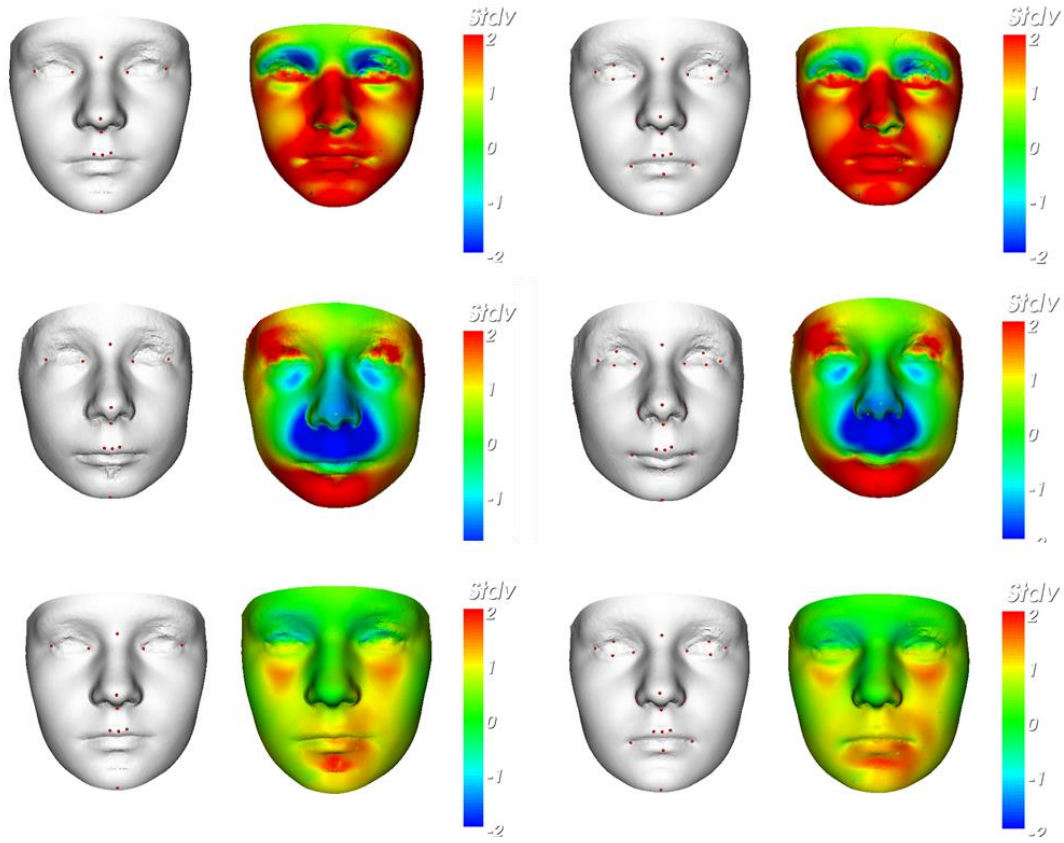


Figure 6.8 Surface synthesis and facial signatures of 3 randomly selected individuals with FAS using models built from point correspondence with (left column) a reduced set of 13 landmark points and (right column) 24 landmarks.

6.5.3 FaceScreen: Clinical User Interaction

Ultimately, the clinician using this software will be responsible for image capture, landmarking, analysis and report generation. Throughout this test phase, we have utilised the Vectra H1 hand-held camera system proving it to be as effective as a static tripod mounted system. Once the three required images have been taken, and the stitched surface generated, it will need to be checked for surface quality by removing the texture from the image revealing the underlying facets. Should any part of the image show artefacts or stitching inconsistencies then the capture process will need to be repeated. An image of suitable quality will then be exported and transferred into the FaceSreen application. On import, the user will be prompted to enter patient information: a unique patient identifier; date of birth; date of scan; ethnic background; and, sex. Immediately, the image will have loaded into in the software in two windows, one with the surface texture and the other with a mean curvature overlay (Figure 6.9). Once loaded, a series of landmarks will have to be manually annotated to the image. We have introduced a visual guide for each landmark to ensure the user is prompted for each placement (see Appendix 6; Figure 9.5). This acts as an indicator of the current landmark in the sequence and an informative reminder of exact localisation. With the addition of left and right exocanthi, subnasale and upper lip-centre points we can calculate an axis to visualise directional curvature (see Appendix 6; Figure 9.6). Adding the inner canthi points for both eyes will allow for the

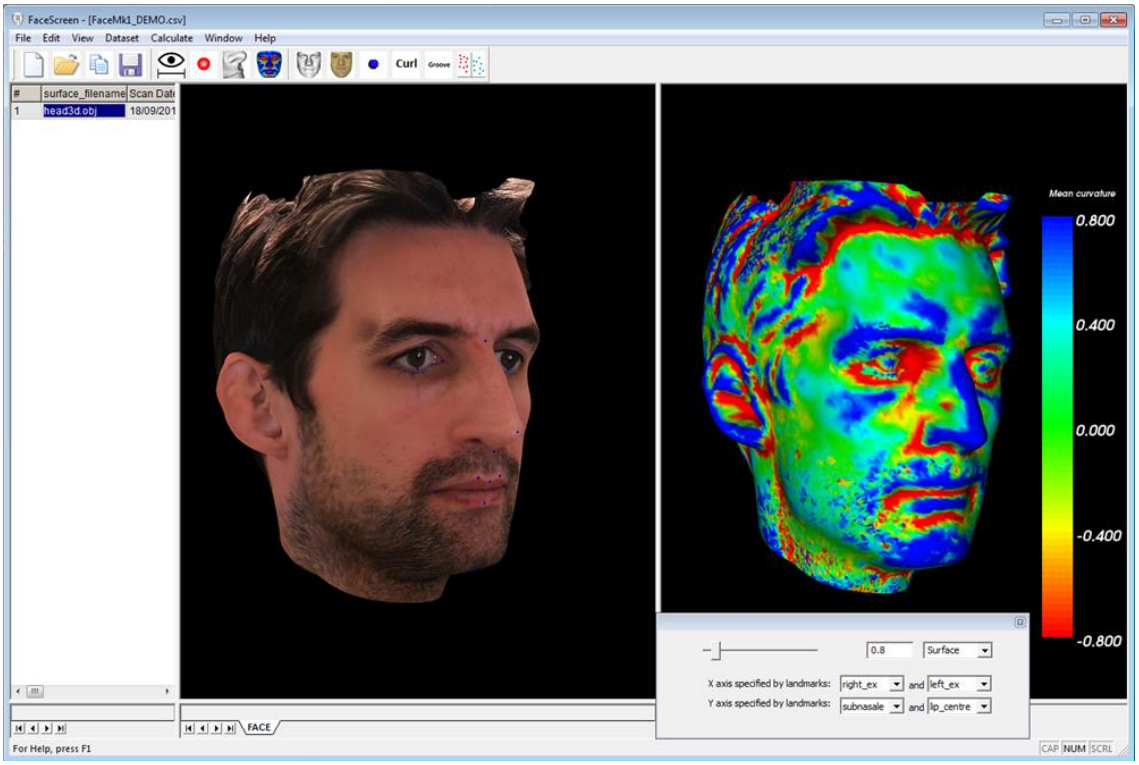


Figure 6.9 FaceScreen user interface showing a loaded image. Left window shows image overlaid with texture, and the right shows the mean curvature heat map overlay.

calculation of PFL, which is displayed in a dialog containing plotted Z-score and percentile rank using normative data from 3D datasets (Figure 6.10).

An individual can be projected into a DSM classification model to determine closeness to either FAS or control group means. The aforementioned discrimination techniques utilising PC values for closest mean classification testing can be applied to the whole face or a particular selected region. Testing will only execute if landmarks in the classification models match that of the individual, so further discrimination testing requires the remaining landmarks to be added by the user. Once added, the user can select from a list of facial regions to obtain an individual classification outcome. For example, if the user wanted to know how closely the philtrum of a subject matches that of the FAS mean compared to the control mean (Figure 6.11).

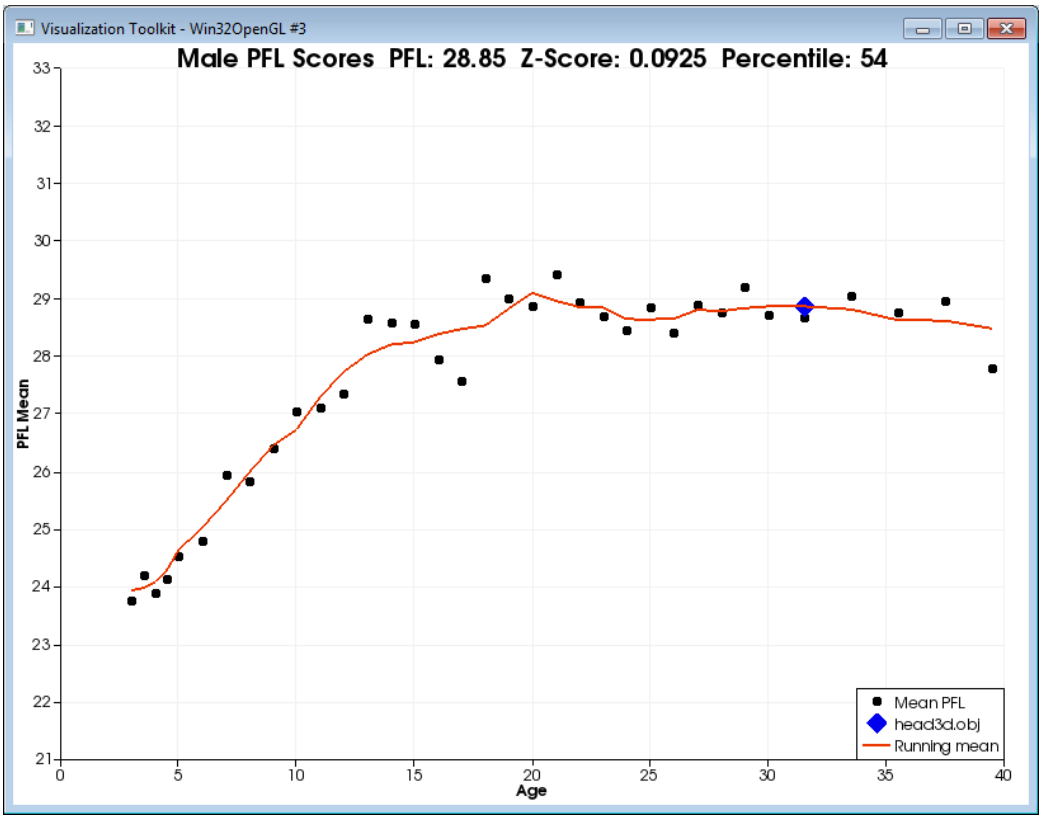


Figure 6.10 Individual PFL plotted with blue diamond, against mean values for age ranges from FaceBase normative data. PFL, z-score and percentile rank are all displayed for the user.

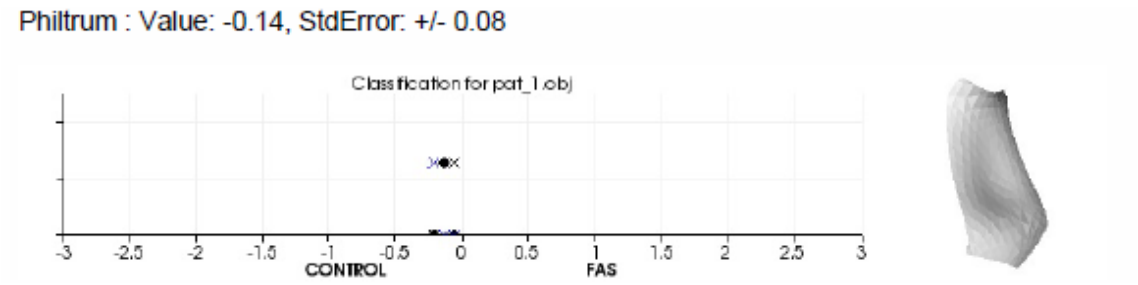


Figure 6.11 Output dialog for discrimination testing for the philtrum of an individual tested using closest mean (control or FAS) classification testing.

6.5.4 Clinical Report Generation and Validation

10 patients seen at the UK FASD clinic underwent the clinical evaluation process described in Figure 6.1, with the addition of a 3D image taken in initial session on day 1. Each image was landmarked using a reduced set of 13 anatomical points (see 6.5.2), and reports were automatically generated to a portable document format (PDF) file summarising the 3D analysis. Each report contains

- patient information such as age, sex, date of birth, date of scan, ethnicity, plus any user-defined attribute
- a screenshot showing profile and portrait views of the subject
- facial signature heat map
- plotted PFL measurement, Z-scored, with percentile rank calculated from age-sex matched means obtained from the FaceBase normative measures database.
- ¾ view screenshots of philtrum without surface texture and with an overlaid curvature heat map showing groove curvature
- control-FAS discrimination test results for the full face and each of the 7 facial regions including nose, malar, midline profile, mandible, eyes, philtrum and upper-lip vermillion

An anonymised example of a report can be seen in Appendix 7.

Ten 3D facial analysis reports generated by FaceScreen were independently reviewed by Dr Mukherjee, alongside the reports generated by the FAS Facial Photographic Analysis Software using 2D photographs. Clinical diagnoses had not yet been assigned to these patients so direct validation in concordance with diagnostic labelling was not possible. However, of the ten patients, three presented facial dysmorphism satisfying FAS diagnostic criteria (detected by the 2D software), confirmed alcohol exposure, and behavioural/cognitive deficits. FaceScreen 3D reports showed these individuals to have a PFL <10th percentile, curvature on the philtrum in fitting with smoothness, and control-FAS classification results scoring positive for FAS in each region of the face. The majority (n=5) of the remaining cases showed a variation of dysmorphic features. Using facial signatures, measurements, curvature analysis and facial region classification we confirmed the presence of: hypoplastic midface; retrognathia; a short button like nose; reduced PFL; and, philtrial smoothness. Two cases did not present any FAS-like facial features of show the presence of any recognised minor facial anomalies. Maternal exposure information for these two cases confirmed drug use during pregnancy, with alcohol use strongly suspected. To conclude, the case by case review and discussion with Dr Mukherjee indicated that FaceScreen 3D reports provided a clinically useful objective facial assessment. Feedback was positive, and minor issues regarding report content and design were noted for the next iteration.

6.6 Discussion

In this chapter, we have implemented the initial development phase of a screening tool for the recognition of facial features associated with PAE using 3D imaging. Techniques included were proven effective for the identification of FASD features based on findings from the previous chapters, and published studies (Suttie *et al.*, 2013, 2017). Working closely with clinicians at the UK FASD clinic, we have preliminarily validated the resulting implementation ‘FaceScreen’,

testing the agreement of our software in line with clinical judgement and reports generated using the FAS Facial 2D Photographic Analysis. After adding a small number of landmarks to a 3D image, FaceScreen, in its current state, provides a set of tools for objective identification of features utilising a combination of normalised heat maps, curvature analysis, closest mean classification testing and linear measurements.

For usability, we implemented a visual landmark guide showing the placement of each point on an example surface. Landmarks were placed in a set sequence for user familiarity, and after minimal training using this process, clinicians were able to annotate a small test set of images with landmarks accurately. We were able to utilise a reduced set of landmarks (13) that encompassed relevant linear measurements, while still ensuring an accurate synthesis for discrimination testing. Using this reduced set of 13 landmarks revealed similar control-FAS discrimination accuracies compared to that of 24 landmark models. The method for PFL measurement from points manually placed on 3D images tends to overestimate the length compared to ruler measurements, and hence, under-diagnose short palpebral fissures. To determine percentile rank for an individual's ruler measured PFL, clinicians use normative data from ruler measurements. The same normative values cannot be used for 3D PFL measurements given the disparity between 3D measurements and ruler based estimates we observed in multiple cohorts. Instead, we utilise age-sex matched means produced from 3D images provided by the FaceBase normative measures database. No significant differences were found between normative values and those from the Caucasian cohort used in this thesis. Unfortunately, at the time of writing only Caucasian data has been made available from this source, so to cover additional ethnic cohorts we must source ethnicity matched measurements.

The updated clinical guidelines (Hoyme *et al.*, 2016) present an analysis scoring minor facial anomalies in 370 (5-8yrs) Caucasian children with a FAS diagnosis based on their observed frequency. The cardinal features; head circumference (≤10th percentile), short PFL (≤10th percentile), smooth philtrum and thin vermillion (4 or 5 on Likert scale) were assigned a score of 3, while other features observed in ≥100 of the subjects were assigned a score of 2. The remaining features observed in fewer than 100 subjects were assigned a score of 1. Minor facial anomalies scoring a 2 consisted of a hypoplastic midface (malar flattening), epicanthal folds, prognathism, decreased interpupillary and inner canthi distance (≤25th percentile), long philtrum (≤90th percentile), anteverted nares and a flat nasal bridge. The dysmorphology assessment to recognise minor anomalies is necessarily subjective, and some can be identified as a binary observation. The presence and degree of features such as midfacial hypoplasia and nasal bridge flattening would be more accurately assessed using imaging-based methods. Results of facial analysis in Chapters 3 and 4 show the importance of evaluating the facial gestalt of FASDs beyond the three cardinal features. In both Cape Coloured and Caucasian populations we identified individuals with heavy prenatal alcohol exposure, without FAS or PFAS criteria, who had an affinity to FAS by their facial dysmorphism. Recognisable features in these individuals included malar flattening, flat nasal bridge, micrognathia and retrognathia. Thus, in this screening tool, we have included the ability to recognise individual dysmorphism in these regions using localised discrimination testing and facial signature heat maps.

No imaging tools exist for estimating the degree of malar/midfacial hypoplasia. By utilising a DSM based approach on only the malar region, we can determine whether this region in an individual is closer to that of the average FAS or average control. To further optimise this process, DSM classifiers could use training sets built utilising only individuals who present with

malar flattening. This way we can determine two means to classify on, one made from a subset of prenatal alcohol-exposed individuals with malar flattening and the other calculated from controls who have normal malar development. Authors of the 2D imaging screening protocol proposed a Likert scale for identifying the degree of midfacial hypoplasia in an early publication (Astley & Clarren 1995; Figure 6.12). This scale, referenced by three 2D profile images of individuals who score either 0 (well-developed midface), 1 ('somewhat developed midface') and 2 ('definitely Hypoplastic midface'). Challenges also arise when trying to associate a hypoplastic midface specifically with prenatal alcohol exposure, as it is a normal characteristic in some ethnicities. This stresses the importance of comparing an individual using ethnically-matched comparison data.

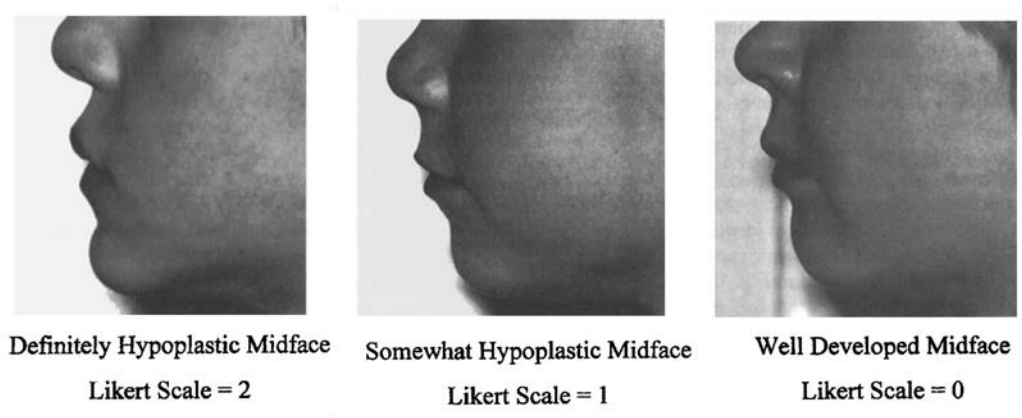


Figure 6.12 Extract from Astley & Clarren 1995. 3-point Likert scale proposed for the identification of hypoplastic midface.

Limitations

The 3D device requires more care than a conventional 2D camera, especially with less co-operative subjects. The Canfield H1 handheld system requires the capture of three consecutive images, and thus, is sensitive to artefacts caused by changes in facial expression and eye movement between shots. A small number of images taken by clinicians at the UK FASD clinic had to be discarded for such issues, but after some familiarisation with the system image quality was maintained.

The face screening software is relatively easy to use but is likely to require separate interfaces and different screening tools for use in secondary (e.g. general practice, clinical genetics) and tertiary settings (e.g. specialist FASD clinic). Future work should focus on streamlining this process to the greatest extent. The main issue was the requirement for landmark placement, which requires user training and is subject to issues with interrater dependent variability. Although we were able to reduce the number of landmarks without impacting the outcome of the analysis, we feel manual landmark placement should be avoided, or at least semi-automated. There are published algorithms for the automated placement of landmarks on the 3D face, using statistical shape analysis (Shu Liang *et al.*, 2013; de Jong *et al.*, 2016), machine learning (Creusot, Pears and Austin, 2012) and surface curvature (Lippold *et al.*, 2014). Future iterations should focus on the integration of such algorithms in an attempt to provide consistent and accurate landmark placement.

The detection of philtral smoothness using curvature can visually identify philtrum convexity, but this does not quantify smoothness. Producing an accurate, quantifiable measure to assess

philtrum smoothness is key to providing a fully objective clinical examination. The ability to determine a measure on a continuous scale enables greater sensitivity when compared to the finite and categorised scoring system provided by the Lip-Philtrum Guide (Astley and Clarren, 2000). To date, there are no imaging techniques in the literature that addresses this issue by providing such a measure, being a FASD specific facial feature it is unlikely that general imaging expertise would focus on.

In a series of tests, we attempted to quantify raw 3D image data of the philtrum to produce such a measure. We refer to this as a philtrum groove index (PGI), which is based on a semi-autonomous method which calculates the volume of the philtral groove (Figure 6.13). For this calculation; the philtrum groove refers to the region bound by a curvature based upper threshold below or at the subnasale, the upper vermillion border and the philtral pillars. This region is semi-automatically identified using a combination of surface curvature and manually placed anthropometric landmarks. An initial upper seed point is determined using an edge-detection style operation identifying the maximum rate of change in curl curvature along a subnasale to lip-centre path. This pseudo-landmark representation initialises curvature-guided geodesic paths to the lip centre, calculated for both left and right sides of the philtrum. The resulting lateral edges of the computed paths typically coincide with positive groove curvature at the philtral pillars. The concatenation of the left and right paths achieves a continuous boundary encompassing the desired philtrum groove region.

Once the region is demarcated, an identical point set is warped to the surface using the boundary points to form a convex hull. An arbitrary plane is used to define positive and negative voxels by projecting the triangular facets of the philtrum and upper convex hull surfaces onto it. The volume is estimated as the difference between the sum of the positive and negative voxel elements. To produce the final measure, we calculate the log of this volume estimation divided by the philtrum length (subnasale-lip centre distance) such that

$$philtrum\ groove\ index = \ log\left(\frac{Vol}{Philrum\ Length}\right).$$

Unfortunately, this technique failed to quantify the philtrum in the manner we had hypothesised. We found the PGI to be highly sensitive to the varying degree of quality of image data, the vastly heterogeneous and complex nature of the philtrum and the ability to accurately and consistently segment the region of interest.

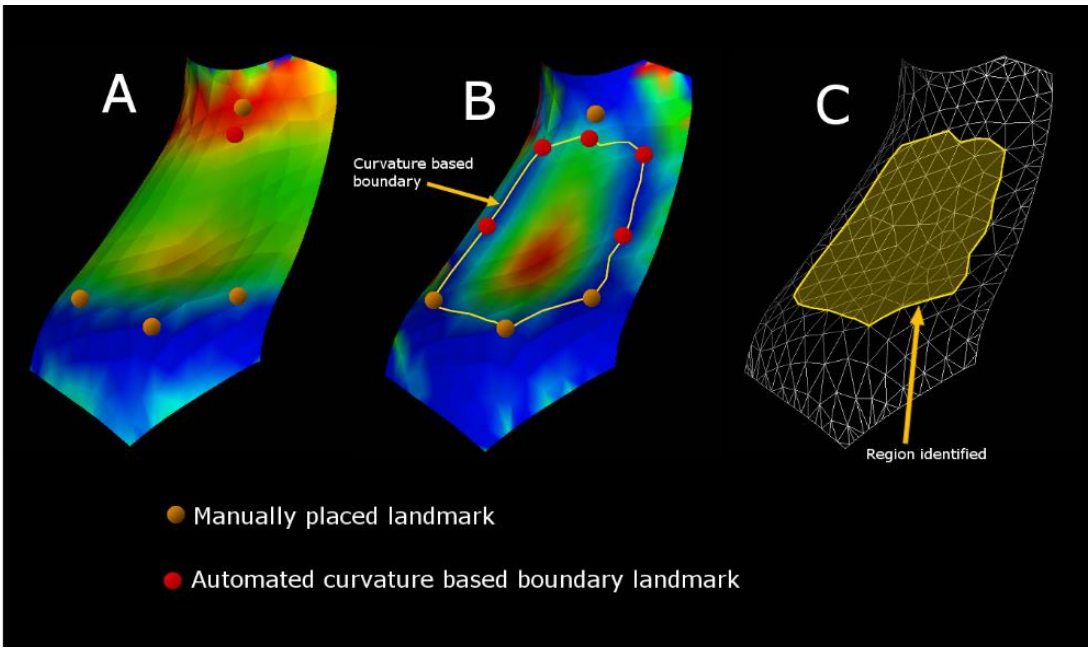


Figure 6.13 An initial seed point is selected by calculating a geodesic path from the subnasale and upper lip centre on to the philtrum surface. The maximum change in gradient of the curl curvature shown in A) along this path from negative (red) to positive (blue/green) defines this point as an upper boundary limit. Left and right upper points are determined by peaks in positive groove curvature (shown blue in B) along lateral paths - typically coincident with the superior section of the philtral pillars. A geodesic path from left and right upper points to the left and right cupid points respectively is calculated. A modified version of Dijkstra's shortest path algorithm in which positive groove curvature influences node selection is used to calculate each path – the resulting paths are typically coincident with the philtral pillars. The lower bounds of the region and final section of the boundary are computed using the shortest path from cupid points to lip centre – typically coincident with the vermillion border of the upper lip. The region boundary discriminated in yellow, outlined in B, and in shaded form in C. The geometry encompassed within this boundary is projected to a plane and thin plate spline warped back using boundary points as reference. This produces a “lid” to the philtrum groove. Individual triangular facets of the contained region and warped geometry of the “lid” are then projected onto an arbitrary plane to produce positive and negative voxels whose net sum is an estimate of the volume of the philtral groove.

6.7 Conclusions

In this chapter, we have investigated the use of incorporating 3D facial analysis into the clinical workflow to assess facial features associated with FASD diagnoses. It is important to note that while this tool can be used to analyse and identify facial features, it alone cannot decipher a diagnosis. To produce a robust method for diagnosis, we need to also consider neurocognitive deficits, behavioural characteristics and other physical anomalies. It would, therefore, be inaccurate and unfitting to assume a tool which solely analyses facial characteristics would be capable of correctly providing a diagnostic label. What this does provide is a series of methods which accurately and objectively identify facial features that could otherwise be misrepresented. The overall result for the clinician allows them to consider a ranked likelihood of the presentation being a consequence of PAE.

Although we have achieved some of the clinical requirements, it is beyond the scope of this project to produce a finished distributable product. However, we have achieved a promising basis for a clinical tool which will be further developed over the coming years as part of the

current CIFASD funded project within the Nuffield Dept. of Obstetrics and Gynaecology, at the University of Oxford.

7 DISCUSSION

In this chapter, we summarise this thesis, provide some further discussion and highlight the contributions and future direction of this work.

7.1 Summary of Findings

We have shown (Chapter 3) that DSM based facial analysis provides a novel approach for the identification of facial effects associated with PAE. This initial analysis addressed some of the core objectives of this thesis, identifying strategies for recognising patterns of facial dysmorphism across the FASD spectrum. We achieved perfect control-FAS discrimination (in a Cape Coloured population) and near perfect agreement with the addition of PFAS, and introduced facial signatures and dynamic morphs as methods of identifying dysmorphism that may otherwise be overlooked. Accurate results were achieved when testing agreement of control-FAS discrimination using only the midline profile, which remained consistently high with the introduction of PFAS. These findings accentuate alcohol's teratogenic effect on midline structures which is apparent from brain imaging studies previously discussed (Chapter 2). In addition to accurately recognising FAS/PFAS differences, we used facial signature graphs to partition the HE group into two distinct groups. Those with a FAS-like facial dysmorphism were identifiable by their position relative to FAS nodes in the graph, while those without clustered separately. Interestingly, when testing neurocognitive performance measures of these two HE subgroups, the FAS-like HEs attained testing scores closer to that of FAS, whereas the remaining were closer to controls. This result suggests FAS-like facial dysmorphism in PAE individuals without FAS/PFAS facial criteria, or clinically identifiable features may exhibit cognitive and behavioural phenotypes closer to that of FAS/PFAS.

Following from the analysis of the Cape Coloured population, we introduce a Caucasian cohort to investigate ethnic differences in the effects of PAE (Chapter 4). In this study, we recapitulated the analysis from chapter 3 and extended the protocol introducing surface curvature analysis. Similar to previous results, accurate discrimination were achieved using relatively small areas of the face. In this analysis, we extended the regions of interest to include malar, mandible and

upper-lip vermillion to cover the full extent of dysmorphia. We found inconsistencies in control-FAS discrimination rates between ethnic cohorts, predominantly in midline facial regions (philtrum, nose, mandible, upper-lip vermillion) and differences in philtrum shape. Curvature analysis was proposed as a means for assisting in the evaluation of PAE associated facial features, and also utilised to explain inconsistencies between the cohorts. Reassuringly, using signature graph analysis for the Caucasian HE group partitioned it into two subsets as with the Cape Coloured cohort. When testing neurocognitive variables, we did not observe the same FAS-like scores for the FAS-like HE group. Additionally, we identified a subset of HE individuals exhibiting significant orbital hypertelorism often associated with DiGeorge syndrome and consistent with mouse studies of exposure timing (Sulik *et al.*, 1986b; Lipinski *et al.*, 2012).

A DSM based technique was introduced to facilitate multi-structural surface analysis applied to investigate relationships between face and brain morphology (Chapter 5). We hypothesised that DSM based analysis could reveal relationships between face-brain and neurocognitive outcomes in PAE groups. For this analysis brain structures were selected known to be disproportionately affected by alcohol teratogenesis. Several studies have identified the presence of severe abnormalities of the corpus callosum in alcohol-exposed individuals. Together with our previous results on midline facial effects provides rationale for its inclusion in this analysis. The caudate nucleus was also selected, as the literature consistently indicates significant volumetric reductions in PAE populations and correlations between volume and neurocognitive measures. Compared to separate face and brain models, selectively combined face region-brain DSMs were able to improve control-FAS discrimination. To understand the dysmorphism present in the caudate nucleus across the PAE population, we utilised the face signature graph protocol previously used for partitioning HEs in the face models. Interestingly, the mean face signature composed of individuals in the FAS-like caudate partition closely resembled that of the mean FAS on each directional axis. There was no distinction between cognitive scores for the FAS-like HE group and the remaining HEs as seen in the Cape Coloured face analysis. Further investigation into caudate nucleus morphology identified asymmetry in controls, with a significant reduction in the FAS group. Furthermore, this reduction in asymmetry correlated with poorer neurocognitive test outcomes in the PAE population (FAS+HE), where no such correlations existed in the control group. This study supports the notion that brain-behaviour relationships exist in structures particularly vulnerable to alcohol teratogenesis.

The final stage of this thesis (Chapter 6) focused on the clinical translation of FASD facial analysis protocols proven to be useful in earlier studies. Working closely with Dr Raja Mukherjee and colleagues at the UK FASD clinic facilitated the development of a facial screening software protoype ‘FaceScreen’. A patient is referred to the UK FASD clinic if prenatal alcohol exposure is strongly suspected/confirmed, and if assessment by clinical geneticists has ruled out differential diagnoses. The patient will undergo a psychological evaluation, neurocognitive tests and a short facial assessment to provide a possible FASD diagnosis. Currently, the facial assessment uses 2D imaging with semi-automated feature recognition combined with a subjective assessment of the lip and philtrum. We proposed that deploying our 3D screening software may accurately and objectively assist facial analysis, particularly in those where features are subtle and more challenging to identify. The integration of facial analysis protocols such as closest mean DSM classification testing, landmark-based PFL measurement, curvature analysis and face signature representations were found to be useful tools for assisting the clinician. Although not a product ready for deployment, we feel the foundations for eventual use as a screening tool have been

laid. Development is ongoing, and full validation will be required in the future to determine viability for clinical integration.

7.2 Further Discussion

Identifying Individuals with HE and the Significance of Minor Anomalies

In the initial face studies (Chapters 3 and 4) we describe a protocol using facial signature graphs of exposed subjects (FAS, PFAS and HE) to select HEs with FAS-like facial dysmorphism. On removing the FAS (and PFAS) from the graph, normalising HEs against themselves, HE individuals remained clustered together. By doing this, we can detect a subset of more severely facially affected individuals, finding expression of specific dysmorphism such as retrognathia and midfacial hypoplasia. However, we are not able to pinpoint an exact homogenous facial profile for these individuals. On closer inspection of the HE subset using individual face signatures, we can identify the prominence of facial anomalies described as 'minor anomalies' in the Hoyme diagnostic criteria (Hoyme et al. 2016). From our findings, it could be suggested that the presence of minor anomalies could play a more significant role in diagnosis. Perhaps the difficulty involved with identifying these features accurately is the reasoning behind their exclusion from diagnostic criteria, which could be overcome with the use of image-based facial analysis.

Other Applications for Combined Dense Surface Models

Chapter 4 introduces combined DSMs to analyse the covariance between face and brain surface morphology. This technique is not limited to face-brain analysis and can be applied to any two structures providing they be represented by a single layer 3D surface. In a concurrent study, combined DSMs were used to for modelling skeletal and face surface shape for Treacher Collins Syndrome (Ibrahim *et al.*, 2016). Combined models utilised the face and underlying skeletal surfaces obtained from CT images, showing close alignment between skeletal and soft tissue models. In this analysis, we were also able to provide an evaluation of surgical outcomes for one patient who had undergone corrective surgery.

Clinical translation

In Chapter 6 of this thesis, we clinically translate techniques proven to be effective for the recognition of PAE associated facial features using 3D images. The development of FaceScreen is ongoing, but the first iteration of the software produced reports providing practical assistance to the clinician. Certain issues remain for practicality and costs associated with 3D camera systems as the current price of the system tested is around £8,000. This camera system also requires a certain level of expertise to use, and parts of the process of image capture to image analysis remain unfriendly to the user. Usability was emphasised as an essential requirement by the clinicians from the UK FASD clinic, and future releases should ensure the protocol is simplified.

At present, using FaceScreen, landmarks are required to be manually placed by the clinician. We were able to provide an image guided procedure for landmarking and prove that fewer landmarks could be used effectively without diminishing control-FAS discrimination accuracy. However, the landmarking process is still prone to inaccuracies and may suffer interrater reliability issues. PFL measurements are particularity sensitive to placement errors as the exocanthi and exocanthi eye landmarks directly alter the measurement outcome. For this

reason, short training sessions for clinicians are required to ensure landmark placement is consistent and in line with those in any classification models an individual would be tested against. However, future iterations of the software should address this by considering alternative methods for dense surface correspondence or the introductions of automated landmarking algorithms.

The Future of 3D Imaging Technology

In this thesis, we used high-end commercially available stereophotogrammetry cameras capable of capturing surface points to submillimetre accuracy. Both the static and tripod mounted cameras are expensive options for image capture, but future technological developments could reduce these costs. In recent years, mobile computing capability has seen dramatic increases in processing power, hardware efficiency and camera technology. The Face Camera Pro is an affordable, hand-held 3D imaging system designed to run on mobile devices from Bellus3D (www.bellus3D.com) capable of capturing photorealistic 3D images of the face. This system attaches to an Android-based tablet or phone and uses an infrared structured-light depth camera to acquire surface geometry at up to 0.4mm resolution with submillimetre accuracy. Additionally, the camera automatically identifies 77 landmark points during capture which are made available to the user. At the time of writing, this camera had not yet been released to the mass market, however, towards the end of this thesis we were able to acquire it on a developer licence for testing. We were able to implement a method for automatically transferring the captured image to a networked laptop, and adapt FaceScreen to read and extract a selection of automated landmark points. The analysis techniques detailed in this thesis (Chapter 6) are then available to the user, and the FASD facial report can be generated instantly with little user interaction. This is a promising, inexpensive (~£800 including the mobile device) and innovative development for 3D facial analysis, but it is not without limitations.

Firstly, the capture process takes 15 seconds, and requires the subject to position themselves correctly and turn their head 180 degrees slowly to the left, then back to the centre, then 180 degrees to the right. Although untested, this may pose an issue for children with severe behavioural problems as they may be unwilling or unable to cooperate. The manufacturer claims to capture images at a geometric resolution of 0.4mm, the Vectra H1 camera in comparison achieves <0.2mm. This difference may seem minimal, but surface shape requirements for FASDs, particularly the philtrum area, requires very high-resolution images. Figure 7.1 shows the difference between the two images in the author's face. Observable differences in the Bellus3D image show there to be a smoothing effect giving a loss of detail (Figure 7.1; col 3), particularly around the philtrum. The Bellus3D camera is still in early stages of development, so this may improve in future iterations. Lastly, the automated landmarking process is regrettably inaccurate. Eye points, in particular, did not perform well, and since the PFL measurement is sensitive to error we require a high level of accuracy for placement. In a test set of ~10 images taken of the same subject, we observed several images inaccurately placed points, and there were inconsistencies in the placement of exocanthi and endocanthi points. Although some limitations exist, this is a promising development for the future of 3D facial analysis, the clinical use of 3D imaging for screening FASDs and other facially affected syndromic conditions.

Figure 7.1 (row 1) The Bellus3D Camera Pro set up integrating with FaceScreen (left) and a self capture using the system (right). Image quality comparison between (row 2) the Vectra H1 hand-held camera (www.canfieldsci.com) and (row 2) the Bellus3D Face Camera Pro (www.bellus3d.com). Col 1 shows the photorealistic 3D image from both systems with surface texture; col 2 shows the underlying 3D surface with a noticeable loss of detail in the Bellus3D image; col 3 shows a close up of the philtrum surface in wireframe view. Mesh triangles in the Vectra H1 camera are dense and uniform whereas in the Bellus3D system they are sparse and non-uniform.

7.3 Future Work

Addressing Differential Diagnoses

The overlap of typical FAS facial features with other genetic and teratogenic conditions that clinicians consider to be differential diagnoses was mentioned but not formally addressed (Chapter 3; Table 3.1). Previous studies using image-based recognition have sought to distinguish between control and FAS populations (Meintjes *et al.*, 2002; Moore *et al.*, 2007; Fang *et al.*, 2008; Mutsvangwa *et al.*, 2010). Few studies have sought to assess accuracy in identifying facial features of these differential conditions comparatively with FASDs. Future work should apply the techniques in this thesis to differential diagnoses, and determine the accuracy of distinguishing them from FASDs. Previous DSM based face analysis has already focused on recognising features of Cornelia de Lange (Bhuiyan *et al.*, 2006), 22q11 deletion and Williams syndromes (Hammond *et al.*, 2005) showing high levels of accuracy discriminating these conditions from controls. These models could form the basis for future studies.

Additional Ethnic Populations

Given the ethnic differences in facial dysmorphism observed in FASD, future work should include additional populations beyond Cape Coloured and Caucasians. Within the CIFASD datasets are a multitude of subjects from a diverse ethnic range including African American, Chinese, Native American and many combinations of mixed ancestry groups. Unfortunately, the numbers of FASD diagnosed subjects in these additional ethnic cohorts were insufficient to include in our analysis. As recruitment continues within CIFASD, we should aim to analyse PAE associated facial features across these additional cohorts using the protocols defined in this thesis.

Further Inclusion of Brain Components

The analysis of face-brain morphology and neurocognitive correlates focused on two brain regions particularly susceptible to alcohol teratogenesis. Further work should focus on recapitulating some of this analysis using additional brain components, as several other candidate regions could be analysed. Greater emphasis should be placed on the ability to include cognitive variables and fully understand intrinsic relationships between face, brain and cognition. However, DSM analysis is not always suitable for brain components given the requirement for landmarking and the complexity of structures which may not exhibit reliable anatomical points of reference. Focus should also be placed on integrating point correspondence techniques which do not require landmarking, extending the application of DSMs.

Utilising Midline Defects

One particular finding, consistent in our analysis, was a recognition of the prominence of midline defects in FASDs observable in the face and brain. Control-FAS discrimination testing using only a midline profile of the face was able to achieve near perfect results. This prompts the possibility of utilising 2D images of the facial profile where 3D is unavailable. By taking a profile image of a subject, the profile can be extracted using generic imaging protocols before being exported into a format suitable for DSM analysis. We will focus some attention on 2D profile images for analysis and potentially incorporate these into the screening process.

Landmarking

The production of DSMs currently requires manual landmark placement before registration of images in anticipation of the generation of a dense correspondence of surface points. Also, this is subject to human error and inter-operator reliability issues, and can also prove challenging when modelling structures difficult to demarcate such as brain components. Methods have been developed for automated landmark localisation (Creusot et al., 2010; Segundo et al., 2007) applicable only to human 3D facial images. A more robust approach is required to enable registration of specific components while maintaining the ability to work indiscriminately with a range of 3D structures. Removal of the landmarking process will prove to be a challenging but revolutionary enhancement to DSM construction. Such improvements will further enhance the ultimate aim of automating the process from image capture to delineation of facial dysmorphism with induced classification to a category of foetal alcohol exposure.

Further Development of FaceScreen

FaceScreen, the software developed in this thesis for FASD screening is currently in its infancy. The initial release of FaceScreen has proven to be a promising start for an objective assessment of the face in FASDs. Although we have proven the ability to distinguish facial features of FASDs from controls accurately, diagnostic categorisation cannot be achieved with the face alone. Other factors would need to be considered to assign a diagnosis, and significant further development is required. Over the course of the CIFASD project, we will work closely with the clinicians in the CIFASD consortium and in UK clinics to determine the best approach and provide a usable and distributable tool to assist in the screening process.

.

8 REFERENCES

Archibald, S. L. *et al.* (2001) 'Brain dysmorphology in individuals with severe prenatal alcohol exposure.', *Developmental medicine and child neurology*, 43(3), pp. 148–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11263683.

Astley, S. J. (2006) 'Comparison of the 4-digit diagnostic code and the Hoyme diagnostic guidelines for fetal alcohol spectrum disorders', *Pediatrics*, 118(4), pp. 1532–1545. doi: 118/4/1532 [pii]\r10.1542/peds.2006-0577.

Astley, S. J. *et al.* (2009) 'Magnetic resonance imaging outcomes from a comprehensive magnetic resonance study of children with fetal alcohol spectrum disorders.', *Alcoholism, clinical and experimental research*, 33(10), pp. 1671–89. doi: 10.1111/j.1530-0277.2009.01004.x.

Astley, S. J. (2015) 'Palpebral fissure length measurement: Accuracy of the fas facial photographic analysis software and inaccuracy of the ruler', *Journal of Population Therapeutics and Clinical Pharmacology*, 22(1), pp. e9–e26.

Astley, S. J. and Clarren, S. K. (1995) 'A fetal alcohol syndrome screening tool.', *Alcoholism, clinical and experimental research*, 19(6), pp. 1565–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12410204.

Astley, S. J. and Clarren, S. K. (1996) 'A case definition and photographic screening tool for the facial phenotype of fetal alcohol syndrome.', *The Journal of pediatrics*, 129(1), pp. 33–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8757560.

Astley, S. J. and Clarren, S. K. (2000) 'Diagnosing the full spectrum of fetal alcohol-exposed individuals: introducing the 4-digit diagnostic code.', *Alcohol and alcoholism (Oxford, Oxfordshire)*, 35(4), pp. 400–410. doi: 10.1093/alcalc/35.4.400.

Astley, S. J. and Clarren, S. K. (2001) 'Measuring the facial phenotype of individuals with prenatal alcohol exposure: correlations with brain dysfunction.', *Alcohol and alcoholism (Oxford, Oxfordshire)*, 36(2), pp. 147–59. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11259212.

Avner, M. *et al.* (2014) 'Validation of the facial photographic method in fetal alcohol spectrum disorder screening and diagnosis', *Journal of Population Therapeutics and Clinical Pharmacology*, 21(1), pp. 106–113.

Aynechi, N. *et al.* (2011) 'Accuracy and precision of a 3D anthropometric facial analysis with and without landmark labeling before image acquisition', *The Angle Orthodontist*, 81(2), pp. 245–252. doi: 10.2319/041810-210.1.

Bhuiyan, Z. A. *et al.* (2006) 'Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience.', *Journal of medical genetics*, 43(7), pp. 568–75. doi: 10.1136/jmg.2005.038240.

BOLAM, J. P. *et al.* (2000) 'Synaptic organisation of the basal ganglia', *Journal of Anatomy*, 196(4), pp. 527–542. doi: 10.1046/j.1469-7580.2000.19640527.x.

Bookstein, F. L. (1997) 'Shape and the Information in Medical Images: A Decade of the Morphometric Synthesis', *Computer Vision and Image Understanding*, 66(2), pp. 97–118. doi: 10.1006/cviu.1997.0607.

Bookstein, F. L. *et al.* (2002) 'Midline corpus callosum is a neuroanatomical focus of fetal alcohol damage.', *The Anatomical record*, 269(3), pp. 162–74. doi: 10.1002/ar.10110.

Bookstein, F. L. *et al.* (2007) 'Many infants prenatally exposed to high levels of alcohol show one particular anomaly of the corpus callosum.', *Alcoholism, clinical and experimental research*, 31(5), pp. 868–79. doi: 10.1111/j.1530-0277.2007.00367.x.

Carter, R. C. *et al.* (2016) 'Fetal Alcohol Growth Restriction and Cognitive Impairment', *Pediatrics*, 138(2), pp. e20160775–e20160775. doi: 10.1542/peds.2016-0775.

Chasnoff, I. J., Wells, A. M. and King, L. (2015) 'Misdiagnosis and Missed Diagnoses in Foster and Adopted Children With Prenatal Alcohol Exposure', *Pediatrics*, 135(2), pp. 264–270. doi: 10.1542/peds.2014-2171.

Chinthapalli, K. *et al.* (2012) 'Atypical face shape and genomic structural variants in epilepsy ', *Brain* . doi: 10.1093/brain/aws232.

Chudley, A. *et al.* (2005) 'Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis', *Canadian Medical Association Journal*, 172(5 Suppl), pp. S1–S21. doi: 10.1503/cmaj.1040302.

Chudley, A. E. (2008) 'Fetal alcohol spectrum disorder: counting the invisible - mission impossible?', *Archives of Disease in Childhood*, 93(9), pp. 721–722. doi: 10.1136/adc.2008.137109.

Clarren, S. K. and Smith, D. W. (1978a) 'The fetal alcohol syndrome.', *The Lamp*, 35(10), pp. 4–7.

Clarren, S. K. and Smith, D. W. (1978b) 'The fetal alcohol syndrome.', *The Lamp*, 35(10), pp. 4–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/251812.

Coles, C. D. *et al.* (2011) 'Memory and brain volume in adults prenatally exposed to alcohol', *Brain and Cognition*. Elsevier Inc., 75(1), pp. 67–77. doi: 10.1016/j.bandc.2010.08.013.

Coles, C. D. *et al.* (2016) 'A Comparison Among 5 Methods for the Clinical Diagnosis of Fetal Alcohol Spectrum Disorders', *Alcoholism: Clinical and Experimental Research*, 40(5), pp. 1000–1009. doi: 10.1111/acer.13032.

Colin Carter, R. *et al.* (2012) 'Effects of Heavy Prenatal Alcohol Exposure and Iron Deficiency Anemia on Child Growth and Body Composition through Age 9 Years', *Alcoholism: Clinical and Experimental Research*, 36(11), pp. 1973–1982. doi: 10.1111/j.1530-0277.2012.01810.x.

Cootes, T. F. *et al.* (1995) 'Active Shape Models-Their Training and Application', *Computer Vision and Image Understanding*, 61(1), pp. 38–59. doi: 10.1006/cviu.1995.1004.

Cootes, T. F., Edwards, G. J. and Taylor, C. J. (1998) 'Active Appearance Models', *Proc. European Conference on Computer Vision (ICCV)*, 2, pp. 484–498. doi: 10.1109/34.927467.

Cox-Brinkman, J. *et al.* (2007) 'Three-dimensional face shape in Fabry disease', *European Journal of Human Genetics*, 15(5), pp. 535–542. doi: 10.1038/sj.ejhg.5201798.

Cranston, M. *et al.* (2009) 'Concordance of three methods for palpebral fissure length measurement in the assessment of fetal alcohol spectrum disorder', *The Canadian Journal of Clinical Pharmacology*, 16(1), pp. 234–241.

Creusot, C., Pears, N. and Austin, J. (2012) '3D landmark model discovery from a registered set of organic shapes', in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, pp. 57–64. doi: 10.1109/CVPRW.2012.6238915.

Croxford, J. and Viljoen, D. (1999) 'Alcohol consumption by pregnant women in the Western Cape.', *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, 89(9), pp. 962–5. doi: 10.1046/j.1464-5491.1999.00150.x.

Day, N. L. and Richardson, G. A. (2004) 'An analysis of the effects of prenatal alcohol exposure on growth: A teratologic model', *American Journal of Medical Genetics*, 127C(1), pp. 28–34. doi: 10.1002/ajmg.c.30013.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (no date) 'California Verbal Learning Test® — Children's Version (CVLT®–C)', *San Antonio, TX: The Psychological Corporation.*, 1994.

Delis, D. C. *et al.* (1994) 'California Verbal Learning Test®—Children's Version (CVLT®–C)', *San Antonio, TX: The Psychological Corporation.*

Dodge, N. C. *et al.* (2009) 'Prenatal alcohol exposure and interhemispheric transfer of tactile information: Detroit and Cape Town findings.', *Alcoholism, clinical and experimental research*, 33(9), pp. 1628–37. doi: 10.1111/j.1530-0277.2009.00994.x.

Donald, K. A. *et al.* (2015) 'Neuroimaging effects of prenatal alcohol exposure on the developing human brain: a magnetic resonance imaging review', *Acta Neuropsychiatrica*, pp. 1–19. doi: 10.1017/neu.2015.12.

Elliott, C. . (2007) 'Differential Ability Scales (2nd ed.)', *San Antonio, TX: Harcourt Assessment*. doi: 10.1177/0829573507302967.

Fang, S. *et al.* (2008) 'Automated diagnosis of fetal alcohol syndrome using 3D facial image analysis.', *Orthodontics & craniofacial research*, 11(3), pp. 162–71. doi: 10.1111/j.1601-6343.2008.00425.x.

Farag, M. and Terrace, C. (2014) 'Diagnostic issues affecting the epidemiology of fetal alcohol spectrum disorders.', Journal of population therapeutics and clinical pharmacology = Journal de la thérapeutique des populations et de la pharamcologie clinique, 21(1), pp. e153-8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24779084.

Fish, E. W. *et al.* (2016) 'Acute alcohol exposure during neurulation: Behavioral and brain structural consequences in adolescent C57BL/6J mice', *Behavioural Brain Research*, 311, pp. 70–80. doi: 10.1016/j.bbr.2016.05.004.

Fraser, S. L. *et al.* (2012) 'Effects of binge drinking on infant growth and development in an Inuit sample', *Alcohol*. Elsevier Ltd, 46(3), pp. 277–283. doi: 10.1016/j.alcohol.2011.09.028.

Fryer, S. L. *et al.* (2012) 'Caudate Volume Predicts Neurocognitive Performance in Youth with Heavy Prenatal Alcohol Exposure', *Alcoholism: Clinical and Experimental Research*, 36(11), pp. 1932–1941. doi: 10.1111/j.1530-0277.2012.01811.x.

Godin, E. A. *et al.* (2010) 'Magnetic Resonance Microscopy Defines Ethanol-Induced Brain Abnormalities in Prenatal Mice: Effects of Acute Insult on Gestational Day 7', *Alcoholism: Clinical and Experimental Research*, 34(1), pp. 98–111. doi: 10.1111/j.1530-0277.2009.01071.x.

Godin, E. A. *et al.* (2011) 'Ventromedian forebrain dysgenesis follows early prenatal ethanol exposure in mice', *Neurotoxicology and Teratology*, 33(2), pp. 231–239. doi: 10.1016/j.ntt.2010.11.001.

Goodall, C. (1991) 'Procrustes methods in the statistical analysis of shape', *Journal of the Royal Statistical Society*, pp. 285–339. doi: 10.2307/2345744.

Gwilliam, J. R., Cunningham, S. J. and Hutton, T. (2006) 'Reproducibility of soft tissue landmarks on three-dimensional facial scans', *European Journal of Orthodontics*, 28(5), pp. 408–415. doi: 10.1093/ejo/cjl024.

Hammond, P. *et al.* (2005) 'Discriminating power of localized three-dimensional facial morphology.', *American journal of human genetics*, 77(6), pp. 999–1010. doi: 10.1086/498396.

Hammond, P. *et al.* (2008) 'Face-brain asymmetry in autism spectrum disorders.', *Molecular psychiatry*, 13(6), pp. 614–23. doi: 10.1038/mp.2008.18.

Hammond, P., Hannes, F., *et al.* (2012) 'Fine-grained facial phenotype-genotype analysis in Wolf-Hirschhorn syndrome.', *European journal of human genetics : EJHG*. Nature Publishing Group, 20(1), pp. 33–40. doi: 10.1038/ejhg.2011.135.

Hammond, P., Suttie, M., *et al.* (2012) 'The face signature of fibrodysplasia ossificans progressiva', *American Journal of Medical Genetics Part A*, 158A(6), pp. 1368–1380. doi: 10.1002/ajmg.a.35346.

Hammond, P. and Suttie, M. (2012) 'Large-scale objective phenotyping of 3D facial morphology.', *Human mutation*, 33(5), pp. 817–25. doi: 10.1002/humu.22054.

Hennessy, R. J. *et al.* (2005) 'Facial surface analysis by 3D laser scanning and geometric morphometrics in relation to sexual dimorphism in cerebral--craniofacial morphogenesis and cognitive function.', *Journal of anatomy*, 207(3), pp. 283–95. doi: 10.1111/j.1469-7580.2005.00444.x.

Heulens, I. *et al.* (2013a) 'Craniofacial characteristics of fragile X syndrome in mouse and man.', *European journal of human genetics : EJHG*. Nature Publishing Group, 21(8), pp. 816–23. doi: 10.1038/ejhg.2012.265.

Heulens, I. *et al.* (2013b) 'Craniofacial characteristics of fragile X syndrome in mouse and man.', *European journal of human genetics : EJHG*. Nature Publishing Group, 21(8), pp. 816–23. doi: 10.1038/ejhg.2012.265.

Hoyme, H. E. *et al.* (2005) 'A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: clarification of the 1996 institute of medicine criteria.', *Pediatrics*, 115(1), pp. 39–47. doi: 10.1542/peds.2004-0259.

Hoyme, H. E. *et al.* (2016) 'Updated Guidelines for Fetal Alcohol Spectrum Disorders', *Pediatrics*, 138(2). Available at: http://pediatrics.aappublications.org/content/138/2/e20154256..info.

Hutton, T. J. (2004) 'Dense Surface Models of the Human Face', *Analysis*.

Hynd, G. W. *et al.* (1993) 'Attention Deficit- Hyperactivity Disorder and Asymmetry of the Caudate Nucleus', *Journal of Child Neurology*, 8(4), pp. 339–347. doi: 10.1177/088307389300800409.

Ibrahim, A. *et al.* (2016) 'Combined soft and skeletal tissue modelling of normal and dysmorphic midface postnatal development', *Journal of Cranio-Maxillofacial Surgery*, 44(11), pp. 1777–1785. doi: 10.1016/j.jcms.2016.08.020.

Jacobson, S. W. *et al.* (2011) 'Impaired Delay and Trace Eyeblink Conditioning in School-Age Children With Fetal Alcohol Syndrome', *Alcoholism: Clinical and Experimental Research*, 35(2), pp. 250–264. doi: 10.1111/j.1530-0277.2010.01341.x.

Jacobson, S. W. *et al.* (2017) 'Heavy Prenatal Alcohol Exposure is Related to Smaller Corpus Callosum in Newborn MRI Scans', *Alcoholism: Clinical and Experimental Research*, 41(5), pp. 965–975. doi: 10.1111/acer.13363.

Janzen, L. A., Nanson, J. L. and Block, G. W. (1995a) 'Neuropsychological evaluation of preschoolers with fetal alcohol syndrome.', *Neurotoxicology and teratology*, 17(3), pp. 273–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7623737.

Janzen, L. A., Nanson, J. L. and Block, G. W. (1995b) 'Neuropsychological evaluation of preschoolers with fetal alcohol syndrome.', *Neurotoxicology and teratology*, 17(3), pp. 273–9.

Jenkinson, M. *et al.* (2012) 'FSL.', *NeuroImage*, 62(2), pp. 782–90. doi: 10.1016/j.neuroimage.2011.09.015.

Jones, K. L. *et al.* (2006) 'Accuracy of the diagnosis of physical features of fetal alcohol syndrome by pediatricians after specialized training.', *Pediatrics*, 118(6), pp. e1734–e1738. doi: 10.1542/peds.2006-1037.

Jones, K. L. and Smith, D. W. (1973) 'RECOGNITION OF THE FETAL ALCOHOL SYNDROME IN EARLY INFANCY Case-reports', pp. 999–1001.

de Jong, M. A. *et al.* (2016) 'An Automatic 3D Facial Landmarking Algorithm Using 2D Gabor Wavelets', *IEEE Transactions on Image Processing*, 25(2), pp. 580–588. doi: 10.1109/TIP.2015.2496183.

Jorge Cardoso, M. *et al.* (2013) 'STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation.', *Medical image analysis*. Elsevier B.V., 17(6), pp. 671–84. doi: 10.1016/j.media.2013.02.006.

Joseph, J. *et al.* (2014) 'Three-dimensional surface deformation-based shape analysis of hippocampus and caudate nucleus in children with fetal alcohol spectrum disorders.', *Human brain mapping*, 35(2), pp. 659–72. doi: 10.1002/hbm.22209.

Kaemingk, K. (2003) 'Learning following prenatal alcohol exposure: performance on verbal and visual multitrial tasks', *Archives of Clinical Neuropsychology*, 18(1), pp. 33–47. doi: 10.1016/S0887-6177(01)00182-2.

Klingenberg, C. P. *et al.* (2010) 'Prenatal alcohol exposure alters the patterns of facial asymmetry.', *Alcohol (Fayetteville, N.Y.)*, 44. doi: 10.1016/j.alcohol.2009.10.016.

Klingenberg, C. P., Barluenga, M. and Meyer, A. (2002) 'Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry', *Evolution*, 56(10), pp. 1909–1920. doi: 10.1111/j.0014-3820.2002.tb00117.x.

Kodituwakku, P. W. *et al.* (1995) 'Specific impairments in self-regulation in children exposed to alcohol prenatally.', *Alcoholism, clinical and experimental research*, 19(6), pp. 1558–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8749827.

Lebel, C., Roussotte, F. and Sowell, E. R. (2011) 'Imaging the impact of prenatal alcohol exposure on the structure of the developing human brain.', *Neuropsychology review*, 21(2), pp. 102–18. doi: 10.1007/s11065-011-9163-0.

Lemoine, P., Harousseau, H. and Borteyru, J. (1968) 'Les enfants de parents alcooliques: anomalies observées', *Ouest Médical*, (21), pp. 476–482.

Lewis, C. E. *et al.* (2015a) 'Verbal Learning and Memory Impairment in Children with Fetal Alcohol Spectrum Disorders', *Alcoholism: Clinical and Experimental Research*, 39(4), pp. 724–732. doi: 10.1111/acer.12671.

Lewis, C. E. *et al.* (2015b) 'Verbal Learning and Memory Impairment in Children with Fetal Alcohol Spectrum Disorders', *Alcoholism: Clinical and Experimental Research*, 39(4), pp. 724–732. doi: 10.1111/acer.12671.

Lipinski, R. J. *et al.* (2012) 'Ethanol-induced face-brain dysmorphology patterns are correlative and exposure-stage dependent.', *PloS one*, 7(8), p. e43067. doi: 10.1371/journal.pone.0043067.

Lippold, C. *et al.* (2014) 'Facial landmark localization by curvature maps and profile analysis', *Head & Face Medicine*, 10(1), p. 54. doi: 10.1186/1746-160X-10-54.

London, L. (1999) 'The "dop" system, alcohol abuse and social control amongst farm workers in South Africa: a public health challenge.', *Social science & medicine*, 48(10), pp. 1407–1414. doi: 10.1016/s0277-9536(98)00445-6.

Mattson, S. N., Riley, E. P., Sowell, E. R., *et al.* (1996) 'A decrease in the size of the basal ganglia in children with fetal alcohol syndrome.', *Alcoholism, clinical and experimental research*, 20(6), pp. 1088–1093. doi: 10.1111/j.1530-0277.1996.tb01951.x.

Mattson, S. N., Riley, E. P., Delis, D. C., *et al.* (1996) 'Verbal learning and memory in children with fetal alcohol syndrome.', *Alcoholism, clinical and experimental research*, 20(5), pp. 810–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8865953.

Mattson, S. N. and Riley, E. P. (1998) 'A review of the neurobehavioral deficits in children with fetal alcohol syndrome or prenatal exposure to alcohol.', *Alcoholism, clinical and experimental research*, 22(2), pp. 279–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9581631.

Mattson, S. N., Schoenfeld, A. M. and Riley, E. P. (2001) 'Teratogenic effects of alcohol on brain and behavior.', *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism*, 25(3), pp. 185–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11810956.

May, P. a. *et al.* (2015) 'Prevalence and characteristics of fetal alcohol syndrome and partial fetal alcohol syndrome in a Rocky Mountain Region City', *Drug and Alcohol Dependence*. Elsevier Ireland Ltd, 155, pp. 118–127. doi: 10.1016/j.drugalcdep.2015.08.006.

May, P. A. *et al.* (2005) 'Maternal risk factors for fetal alcohol syndrome in the Western Cape Province of South Africa: A population-based study', *American Journal of Public Health*, 95(7), pp. 1190–1199. doi: 10.2105/AJPH.2003.037093.

May, P. A. *et al.* (2008) 'Maternal Risk Factors for Fetal Alcohol Syndrome and Partial Fetal Alcohol Syndrome in South Africa: A Third Study', *Alcoholism: Clinical and Experimental Research*, 32(5), pp. 738–753. doi: 10.1111/j.1530-0277.2008.00634.x.

May, P. A. *et al.* (2010) 'Population Differences in Dysmorphic Features Among Children With Fetal Alcohol Spectrum Disorders', *Journal of Developmental & Behavioral Pediatrics*, 31(4), pp. 304–316. doi: 10.1097/DBP.0b013e3181dae243.

May, P. A. *et al.* (2014) 'Dietary intake, nutrition, and fetal alcohol spectrum disorders in the Western Cape Province of South Africa', *Reproductive Toxicology*, 46, pp. 31–39. doi: 10.1016/j.reprotox.2014.02.002.

May, P. A. *et al.* (2016) 'Breastfeeding and maternal alcohol use: Prevalence and effects on child outcomes and fetal alcohol spectrum disorders', *Reproductive Toxicology*. Elsevier Inc., 63, pp. 13–21. doi: 10.1016/j.reprotox.2016.05.002.

May, P. A. and Gossage, J. P. (2001) 'Estimating the prevalence of fetal alcohol syndrome. A summary.', *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism*, 25(3), pp. 159–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11810953.

May, P. a *et al.* (2000) 'Epidemiology of fetal alcohol syndrome in a South African community in the Western Cape Province.', *American journal of public health*, 90(12), pp. 1905–12. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1446431&tool=pmcentrez&rendertype=abstract.

McGee, C. L. (2005) 'Fetal alcohol spectrum disorders: an overview with emphasis on changes in brain and behavior.', *Experimental biology and medicine (Maywood, N.J.)*, 230(6), pp. 357–365. Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15956765.

Meintjes, E. M. *et al.* (2002) 'A stereo-photogrammetric method to measure the facial dysmorphology of children in the diagnosis of fetal alcohol syndrome.', *Medical engineering & physics*, 24(10), pp. 683–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12460727.

Meintjes, E. M. *et al.* (2010) 'An fMRI study of number processing in children with fetal alcohol syndrome', *Alcoholism: Clinical and Experimental Research*, 34(8), pp. 1450–1464. doi: 10.1111/j.1530-0277.2010.01230.x.

Modat, M. *et al.* (2010) 'Fast free-form deformation using graphics processing units', *Computer Methods and Programs in Biomedicine*, 98(3), pp. 278–284. doi: 10.1016/j.cmpb.2009.09.002.

Moore, E. S. *et al.* (2007) 'Unique facial features distinguish fetal alcohol syndrome patients and controls in diverse ethnic populations.', *Alcoholism, clinical and experimental research*, 31(10), pp. 1707–13. doi: 10.1111/j.1530-0277.2007.00472.x.

Mutsvangwa, T. E. M. *et al.* (2010) 'Morphometric analysis and classification of the facial phenotype associated with fetal alcohol syndrome in 5- and 12-year-old children.', *American journal of medical genetics. Part A*, 152A(1), pp. 32–41. doi: 10.1002/ajmg.a.33137.

Myers, B. (2013) 'Barriers to alcohol and other drug treatment use among Black African and Coloured South Africans', *BMC health services research*. BMC Health Services Research, 13(1), p. 177. doi: 10.1186/1472-6963-13-177.

Myers, B. *et al.* (2013) 'Ethnic differences in alcohol and drug use and related sexual risks for HIV among vulnerable women in Cape Town, South Africa: implications for interventions.', *BMC public health*. BMC Public Health, 13(1), p. 174. doi: 10.1186/1471-2458-13-174.

Nardelli, A. *et al.* (2011) 'Extensive deep gray matter volume reductions in children and adolescents with fetal alcohol spectrum disorders.', *Alcoholism, clinical and experimental research*, 35(8), pp. 1404–17. doi: 10.1111/j.1530-0277.2011.01476.x.

Norman, A. L. *et al.* (2009) 'Neuroimaging and fetal alcohol spectrum disorders.', *Developmental disabilities research reviews*, 15(3), pp. 209–17. doi: 10.1002/ddrr.72.

O'Connor, M. J. and Paley, B. (2009) 'Psychiatric conditions associated with prenatal alcohol exposure', *Developmental Disabilities Research Reviews*, 15(3), pp. 225–234. doi: 10.1002/ddrr.74.

O'Dwyer, L. *et al.* (2016) 'Decreased left caudate volume is associated with increased severity of autistic-like symptoms in a cohort of ADHD patients and their unaffected siblings', *PLoS ONE*, 11(11), pp. 1–20. doi: 10.1371/journal.pone.0165620.

O'Leary-Moore, S. K. *et al.* (2011) 'Magnetic resonance-based imaging in animal models of fetal alcohol spectrum disorder.', *Neuropsychology review*, 21(2), pp. 167–85. doi: 10.1007/s11065-011-9164-z.

Oelofse, A. *et al.* (2002) 'Disadvantaged black and coloured infants in two urban communities in the Western Cape, South Africa differ in micronutrient status', *Public Health Nutrition*, 5(02). doi: 10.1079/PHN2002263.

Ourselin, S. *et al.* (2000) 'Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images', in Delp, S. L., DiGoia, A. M., and Jaramaz, B. (eds) *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2000: Third International Conference, Pittsburgh, PA, USA, October 11-14, 2000. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 557–566. doi: 10.1007/978-3-540-40899-4_57.

Parnell, S. E. *et al.* (2014) 'Dysmorphogenic Effects of First Trimester-Equivalent Ethanol Exposure in Mice: A Magnetic Resonance Microscopy-Based Study.', *Alcoholism, clinical and experimental research*, pp. 1–7. doi: 10.1111/acer.12464.

Pfefferbaum, A. *et al.* (2013) 'Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85years) measured with atlas-based parcellation of MRI', *NeuroImage*, 65(1), pp. 176–193. doi: 10.1016/j.neuroimage.2012.10.008.

Phillip Gossage, J. *et al.* (2014) 'Alcohol use, working conditions, job benefits, and the legacy of the "dop" system among farm workers in the Western Cape Province, South Africa: Hope despite high levels of risky drinking', *International Journal of Environmental Research and Public Health*, 11(7), pp. 7406–7424. doi: 10.3390/ijerph110707406.

Popova, S. *et al.* (2017) 'Estimation of national, regional, and global prevalence of alcohol use during pregnancy and fetal alcohol syndrome: a systematic review and meta-analysis', *The Lancet Global Health*. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND license, (17), pp. 1–10. doi: 10.1016/S2214-109X(17)30021-9.

Postema, F. A. *et al.* (2017) 'Validation of a clinical screening instrument for tumour predisposition syndromes in patients with childhood cancer (TuPS): protocol for a prospective,

observational, multicentre study', *BMJ Open*, 7(1), p. e013237. doi: 10.1136/bmjopen-2016-013237.

Rasmussen, C. (2005) 'Executive functioning and working memory in fetal alcohol spectrum disorder.', *Alcoholism, clinical and experimental research*, 29(8), pp. 1359–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16131842.

Riley, E. P. *et al.* (1995a) 'Abnormalities of the corpus callosum in children prenatally exposed to alcohol.', *Alcoholism, clinical and experimental research*, 19(5), pp. 1198–202. doi: http://dx.doi.org/10.1111/j.1530-0277.1995.tb01600.x.

Riley, E. P. *et al.* (1995b) 'Abnormalities of the corpus callosum in children prenatally exposed to alcohol.', *Alcoholism, clinical and experimental research*, 19(5), pp. 1198–202. doi: http://dx.doi.org/10.1111/j.1530-0277.1995.tb01600.x.

Riley, E. P., McGee, C. L. and Sowell, E. R. (2004) 'Teratogenic effects of alcohol: a decade of brain imaging.', *American journal of medical genetics. Part C, Seminars in medical genetics*, 127C(1), pp. 35–41. doi: 10.1002/ajmg.c.30014.

Roebuck, T. M., Mattson, S. N. and Riley, E. P. (1999) 'Behavioral and Psychosocial Profiles of Alcohol-Exposed Children', *Alcoholism: Clinical & Experimental Research*, 23(6), p. 1070. doi: 10.1097/00000374-199906000-00016.

Roozen, S. *et al.* (2016) 'Worldwide Prevalence of Fetal Alcohol Spectrum Disorders: A Systematic Literature Review Including Meta-Analysis', *Alcoholism: Clinical and Experimental Research*, 40(1), pp. 18–32. doi: 10.1111/acer.12939.

Roussotte, F. (2012) 'Regional brain volume reductions relate to facial dysmophology and neurocognitive function in fetal alcohol spectrum disorders', *Human Brain Mapping*, 33(4). doi: 10.1002/hbm.21260.REGIONAL.

Roussotte, F. F. *et al.* (2012) 'Regional brain volume reductions relate to facial dysmorphology and neurocognitive function in fetal alcohol spectrum disorders.', *Human brain mapping*, 33(4), pp. 920–37. doi: 10.1002/hbm.21260.

Schoot, R. A. *et al.* (2017) 'Facial asymmetry in head and neck rhabdomyosarcoma survivors', *Pediatric Blood and Cancer*, 64(10), pp. 1–8. doi: 10.1002/pbc.26508.

Shankar, K. *et al.* (2006) 'Physiologic and genomic analyses of nutrition-ethanol interactions during gestation: Implications for fetal ethanol toxicity.', *Experimental biology and medicine (Maywood, N.J.)*, 231(8), pp. 1379–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16946407.

Shu Liang *et al.* (2013) 'Improved detection of landmarks on 3D human face data', in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 6482–6485. doi: 10.1109/EMBC.2013.6611039.

Sizemore, R. J. *et al.* (2016) 'Viral vector-based tools advance knowledge of basal ganglia anatomy and physiology', *Journal of Neurophysiology*, 115(4), pp. 2124–2146. doi: 10.1152/jn.01131.2015.

Sowell, E. (2002) 'Mapping Cortical Gray Matter Asymmetry Patterns in Adolescents with Heavy Prenatal Alcohol Exposure', *NeuroImage*, 17(4), pp. 1807–1819. doi: 10.1006/nimg.2002.1328.

Sowell, E. R. *et al.* (2001) 'Mapping callosal morphology and cognitive correlates: Effects of heavy prenatal alcohol exposure', *Neurology*, 57(2), pp. 235–244. doi: 10.1212/WNL.57.2.235.

Streissguth, A. P. *et al.* (1991) 'Fetal alcohol syndrome in adolescents and adults.', *JAMA*, 265(15), pp. 1961–7.

Sulik, K. K. *et al.* (1986a) 'Fetal alcohol syndrome and DiGeorge anomaly: critical ethanol exposure periods for craniofacial malformations as illustrated in an animal model.', *American journal of medical genetics. Supplement*, 2, pp. 97–112. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3146306.

Sulik, K. K. *et al.* (1986b) 'Fetal alcohol syndrome and DiGeorge anomaly: critical ethanol exposure periods for craniofacial malformations as illustrated in an animal model.', *American journal of medical genetics. Supplement*, 2, pp. 97–112.

Suttie, M. *et al.* (2013) 'Facial dysmorphism across the fetal alcohol spectrum.', *Pediatrics*, 131(3), pp. e779-88. doi: 10.1542/peds.2012-1371.

Suttie, M. *et al.* (2017) 'Facial Curvature Detects and Explicates Ethnic Differences in Effects of Prenatal Alcohol Exposure', *Alcoholism: Clinical and Experimental Research*, pp. 1–13. doi: 10.1111/acer.13429.

Tzou, C.-H. J. *et al.* (2014) 'Comparison of three-dimensional surface-imaging systems.', *Journal of plastic, reconstructive & aesthetic surgery : JPRAS*, 67(4), pp. 489–97. doi: 10.1016/j.bjps.2014.01.003.

Uecker, A. and Nadel, L. (1996) 'Spatial locations gone awry: Object and spatial memory deficits in children with fetal alcohol syndrome', *Neuropsychologia*, 34(3), pp. 209–223. doi: 10.1016/0028-3932(95)00096-8.

Wattendorf, D. and Muenke, M. (2005) 'Fetal Alcohol Spectrum Disorders', *American Family Physician*, 72(2), pp. 279–85. doi: 10.1007/s10995-011-0844-3.

Weinberg, J., D'Alquen, G. and Bezio, S. (1990) 'Interactive effects of ethanol intake and maternal nutritional status on skeletal development of fetal rats.', *Alcohol (Fayetteville, N.Y.)*, 7(5), pp. 383–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2222841.

Willford, J. *et al.* (2010) 'Caudate asymmetry: A neurobiological marker of moderate prenatal alcohol exposure in young adults', *Neurotoxicology and Teratology*, 32(6), pp. 589–594. doi: 10.1016/j.ntt.2010.06.012.

De Wit, E. *et al.* (2010) 'Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape', *Human Genetics*, 128(2), pp. 145–153. doi: 10.1007/s00439-010-0836-1.

Wozniak, J. R. *et al.* (2011) 'Inter-hemispheric functional connectivity disruption in children with prenatal alcohol exposure.', *Alcoholism, clinical and experimental research*, 35(5), pp. 849–61. doi: 10.1111/j.1530-0277.2010.01415.x.

Wozniak, J. R. *et al.* (2016) 'Functional connectivity abnormalities and associated cognitive deficits in fetal alcohol Spectrum disorders (FASD)', *Brain Imaging and Behavior*. Brain Imaging and Behavior. doi: 10.1007/s11682-016-9624-4.

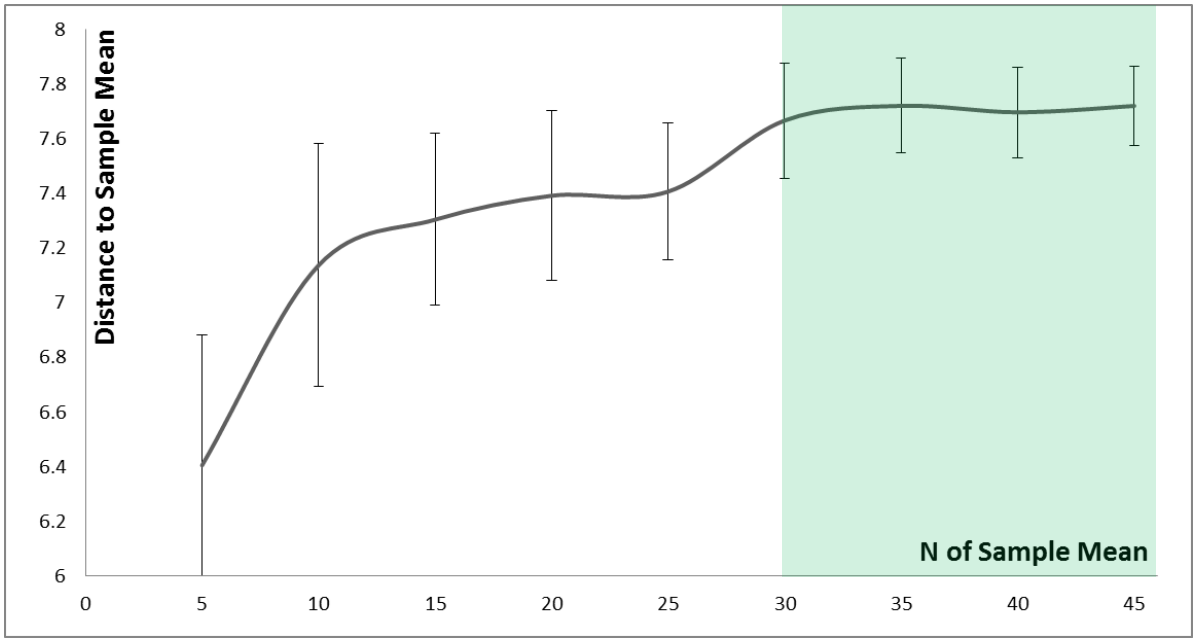
Yang, Y., Roussotte, F., *et al.* (2012) 'Abnormal cortical thickness alterations in fetal alcohol spectrum disorders and their relationships with facial dysmorphology.', *Cerebral cortex (New York, N.Y. : 1991)*, 22(5), pp. 1170–9. doi: 10.1093/cercor/bhr193.

Yang, Y., Phillips, O. R., *et al.* (2012) 'Callosal thickness reductions relate to facial dysmorphology in fetal alcohol spectrum disorders.', *Alcoholism, clinical and experimental research*, 36(5), pp. 798–806. doi: 10.1111/j.1530-0277.2011.01679.x.

9 Appendices

APPENDIX 1 MATCHED MEANS STABILITY

Throughout this thesis, we calculate signature heat maps by normalising point difference using 35 (typically control) surfaces as a reference to calculate matched means. To determine if n=35 is sufficient for matched mean comparisons, we test a series of means (n=5 to n=45, increments of +5) calculated from randomly selected control surfaces from a Cape Coloured face DSM built representing 99% of variance (used in chapters 3 and 4). We calculate the Euclidean distance of PC values from each sample mean, to each surface used to construct it to determine the influence that any one surface has on a generated mean. As the reference sample N increases, so does the distance from each surface used to construct it. Likewise, the sample N increases as the standard error rate decreases indicating less variability of the mean. The plot below shows mean distance values with standard error bars, we observe a plateau around N=30 for the mean distance and standard error - indicting stability of the calculated mean above this level.



APPENDIX 2 COLLAPSED SIGNATURE GRAPHS AND DISPERSION INDICES

Signature graphs presented in Chapter 4 of this thesis are also presented in a simplified 'collapsed' form with an accompanying dispersion index for each subgroup. To achieve this, we apply a graph-theoretic construction referred to as edge contraction, to each subgroup labelled by colour. Figure 4.4 from Chapter 4 shows a collapsed representation of a philtrum signature graph containing the ethnic subgroups: Cape Coloured and Caucasian. This graph is formed by collapsing any edge connecting nodes of the same subgroup (colour), in this case, ethnicity, and replacing the nodes by a single new node labelled by the number of nodes that have been collapsed into it. This process recursively continues until all edges link only nodes of different subgroups. The new nodes representing the collapsed form are labelled by, and illustrated by a radius proportionate to, the number of nodes collapsed to form them. Collapsed signature graphs provide a much-simplified form of the original and provide a visual indication of locally connected signatures which were from the same group. This is particularly useful for identifying the level of homogeneity of a group in terms of face signature variability. A small number of large nodes composed of many collapsed nodes indicates a relatively high level of homogeneity, whereas large numbers of the same grouped nodes of size 1 would suggest the opposite.

We can quantify this using a measure of entropy that we refer to as dispersion. Size labels from sets of collapsed nodes indicate how a particular group has been partitioned. A partition $\{\mathrm{P}_{1,..,}\mathrm{P}_{\mathrm{k}}\}$ of a set S of size n has a dispersion that is quantified as:

$$\mathrm{dispersion}\left(\{P_{1,..,}P_k\}\right) = 1 - \left(\frac{1}{\mathrm{nlog}(n)}\right)\left\{\sum_{i=1}^{k} \mathrm{p}_i \log(\mathrm{p}_i)\right\} \qquad \text{where } \mathrm{p}_i = |\mathrm{P}_i|$$

The dispersion index is between 0 and 1 and in extreme cases for each:

a. if all signatures for group G collapse into a single node so that $\mathrm{k} = 1, \mathrm{P}_1 = \mathrm{G}$ the $\mathrm{dispersion}(\{\mathrm{G}\}) = 0$ indicating complete homogeneity relative to any other group in the graph
b. if no signatures in group G collapse, then each node in the graph will already be connected to a different group, k is equal to all group nodes such that $\mathrm{k} = \mathrm{n}, \mathrm{P}_1 =$ G the $\mathrm{dispersion}\left(\{P_{1,..,}P_n\}\right) = 1$ indicating complete heterogeneity relative to any other group.

APPENDIX 3 CAPE COLOURED AND CAUCASIAN COLLAPSED SIGNATURE GRAPHS

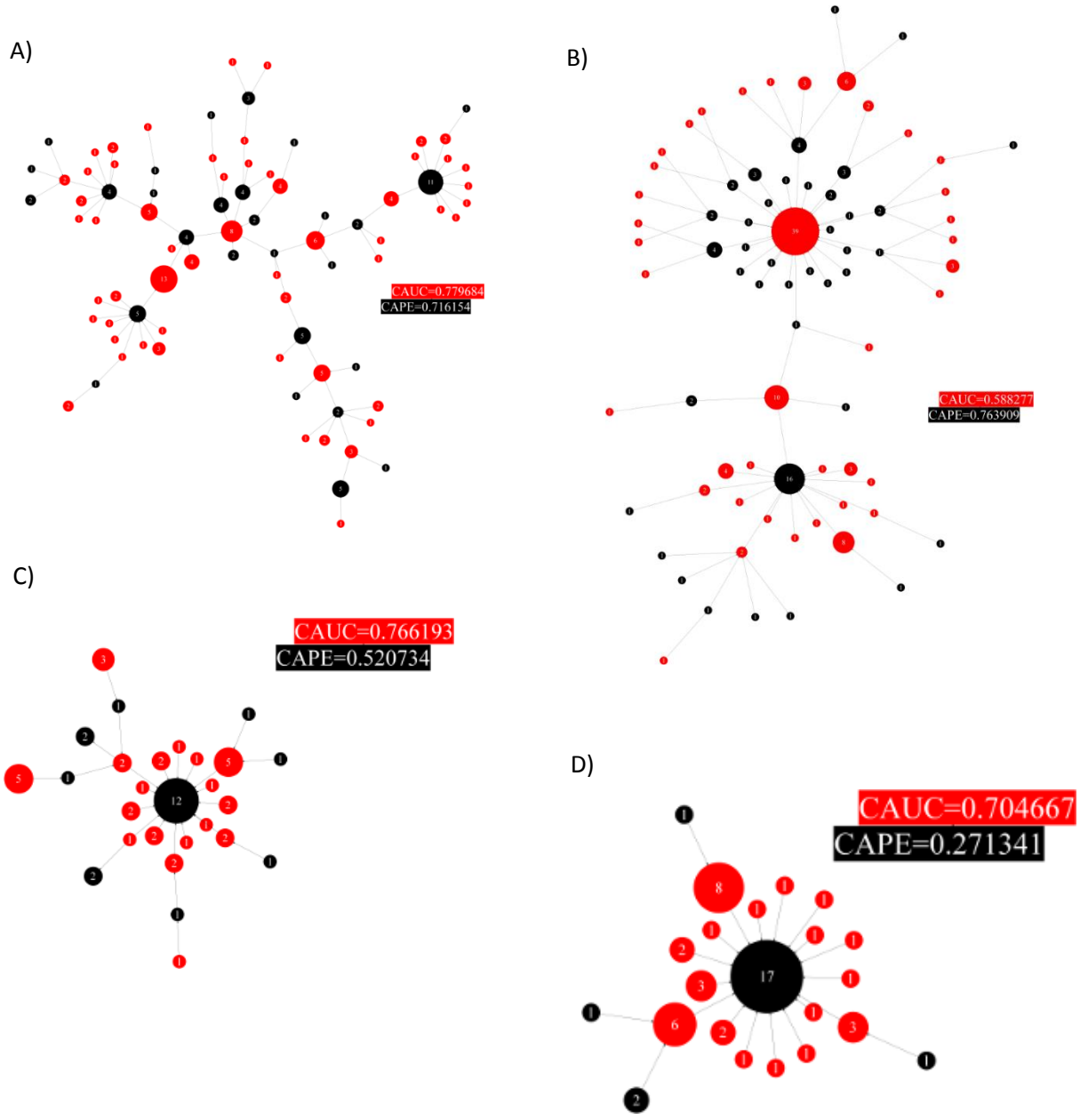


Figure 9.1 Collapsed signature graphs for the face for combined Caucasian and Cape Coloured subsets:

control full face : displacement (A) and curvature (B)

FAS full face : displacement (C) and curvature (D)

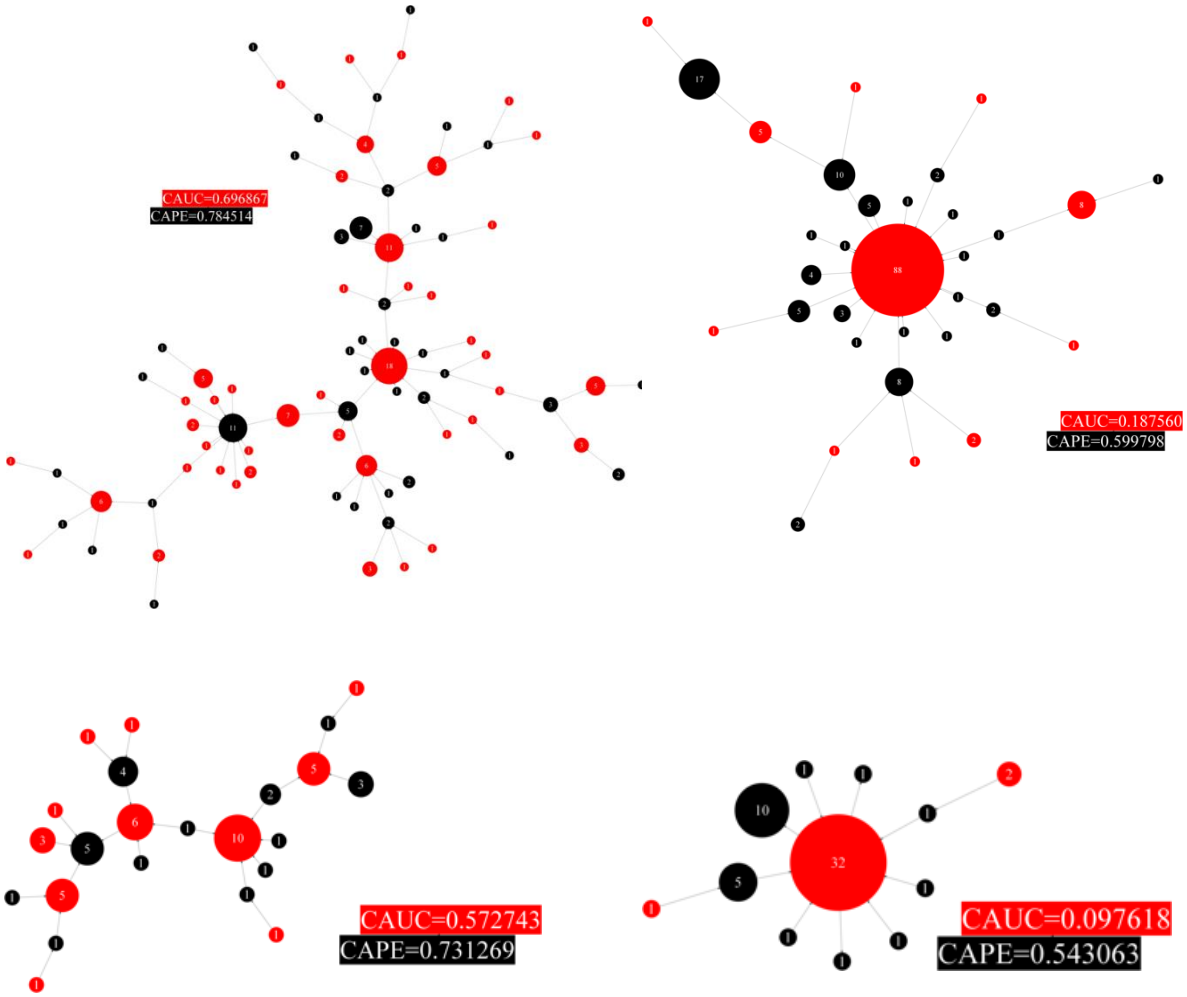


Figure 9.2 Collapsed signature graphs for the philtrum for combined Caucasian and Cape Coloured subsets:

control philtrum: displacement (A) and groove (B)

FAS philtrum: displacement (C) and groove (D)

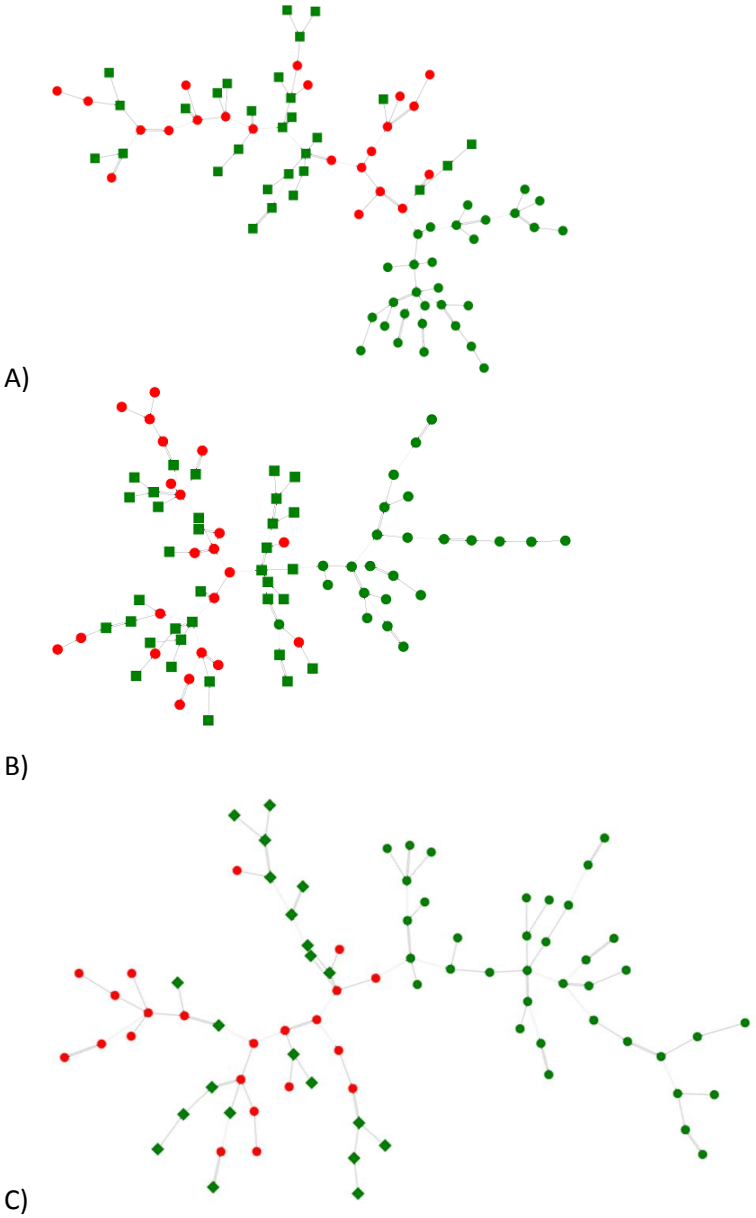


Figure 9.3 Face displacement signature graphs for Cape Coloured exposed (FAS+HE) for the following face patches: A) eyes; B) mandible; C) malar region. In each case, the HE subset is partitioned into a subset clustering with the FAS subset (HE-FAS) and a subset more control-like in its face differences from controls (HE-HC). The colour code scheme uses red for FAS, green squares for HE-FAS and green circles for HE-HC.

APPENDIX 4 COGNITIVE OUTCOME VARIABLES

CVLT Outcome Variables

CVLT-C Test	Description
Discriminability	Ability of the child to discriminate target words from distractors. Number of correct hits subtracted from 15
False Positives (Total)	Items that were incorrectly identified by the child as being on List A when it was not
List A Long Delay Cued Recall	Total correct responses from List A for cued word recall after a 20 minute delay
List A Short Delay Cued Recall	Total correct responses from List A for cued word recall after a 1 minute delay
List A Long Delay Free Recall	Total correct responses recalling as many items from list A as possible in any order after a 20 minute delay
Correct Recognition Hits	Number of correctly identified words

DAS-2 Outcome Variables

DAS-2 Test	Description
GCA Composite Standard Score	Raw score converted to standard age-adjusted score for assessing general cognitive abilities (aka: IQ Score)
Spatial Ability Standard Score	Raw score converted to standard age-adjusted score for assessing capacity to understand and remember spatial relations among objects.
Verbal Similarities Ability Score	Assessing explanations on how 3 concepts or objects go together and what they are.
Recall of Designs Ability	Ability to draw geometric and abstract from memory
Sequential and Quantitative reasoning Ability	Determining mathematical relationships between pairs of numbers and then being able to apply to a second pair of numbers.

APPENDIX 5 SITE DIFFERENCES

	ATL (n=8)	LA (n=37)	MN (n=27)	SD (n=48)	F or X^2	P-value
Control	13.13 ± 1.18	12.33 ± 4.02	12.09 ± 5.2	11.61 ± 1.1	2.82	0.69
FAS	9.69 ± 0.23	10.58 ± 1.50	9.22 ± 1.84	11.11 ± 3.46	3.16	0.25
HE	9.88 ± 0.35	11.51 ± 3.72	11.02 ± 2.99	12.31 ± 3.82	2.80	0.10
ALL	10.6± 1.65	11.69 ± 1.93	11.06 ± 1.57	11.97 ± 2.08	2.07	0.11

Table 9.1 Comparing mean caudate asymmetry indices (± SD) between sites, for each diagnostic category: ATL=Atlanta (n=8); LA=Los Angeles (n=37); Minneapolis (n=27); and, SD=San Diego (n=48), no significant differences exist across sites.

Caudate Nucleus Exposed Signature Graph, Labelled by Site

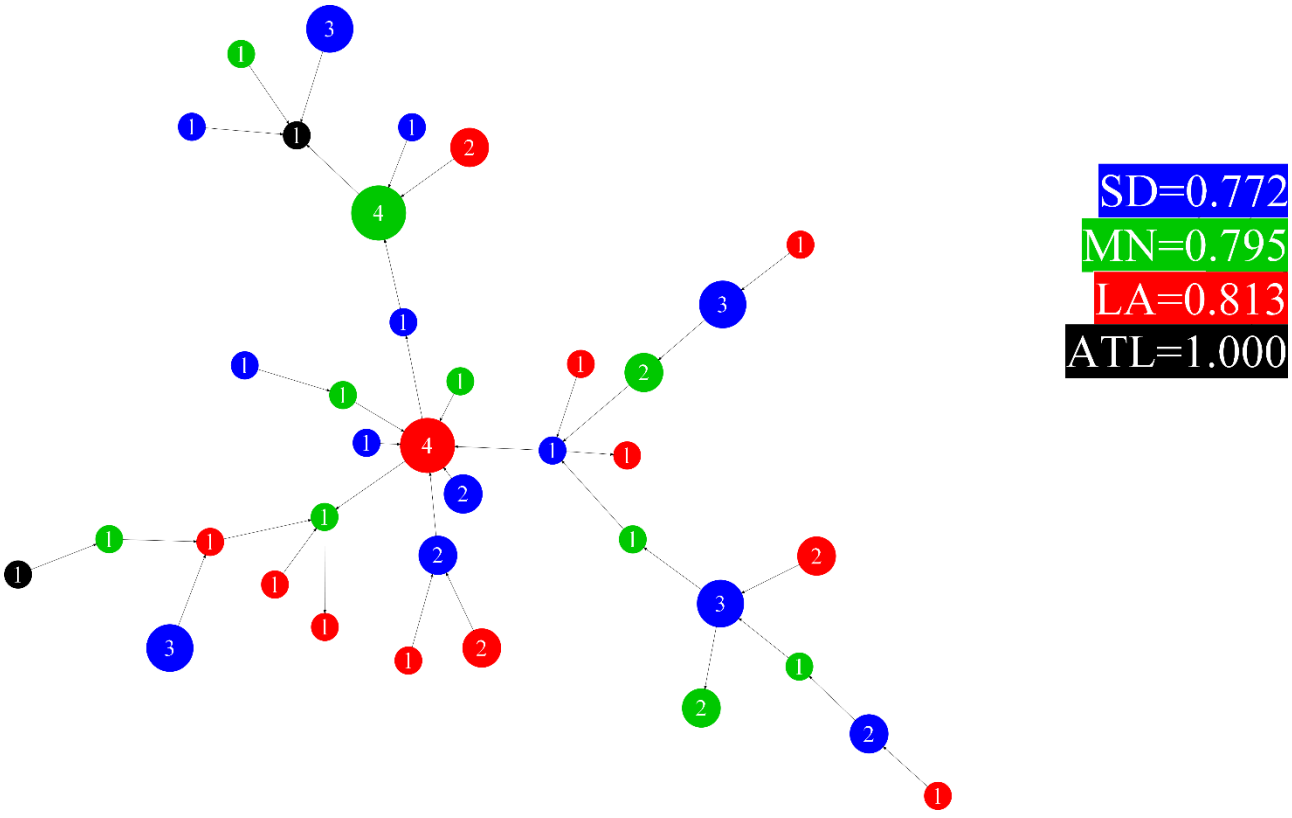


Figure 9.4 Collapsed signature graph of caudate nucleus of exposed individuals (FAS and HE) coloured and grouped by site. For the purpose of determining if there are any site specific differences in caudate dysmorphology that may influence signature connections and grouping. Dispersion indices are quoted for each site (San Diego = 0.772; Minnesota=0.795; Los Angeles=0.813; Atlanta=1.00), where 0 represents compete group affinity and homogeneity and 1 represents complete dispersion with no nodes of the same group are linked.

APPENDIX 6 FACESCREEN: USER WORKFLOW AND INTERFACE

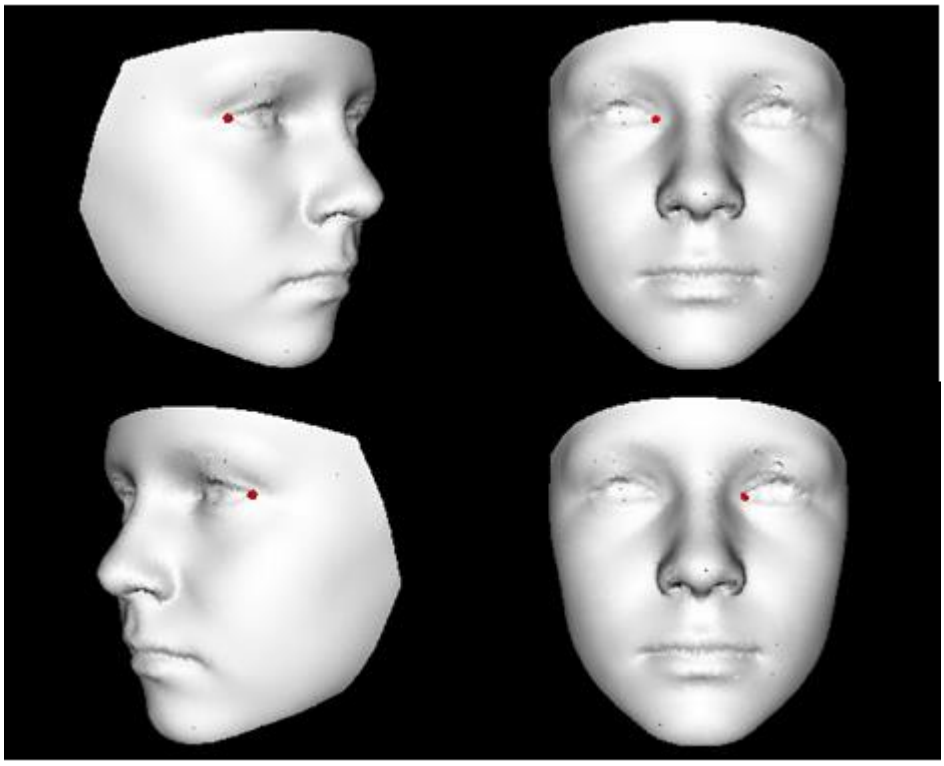


Figure 9.5 Landmark visual placement guide for the 1st 4 points in the sequence right exocanthian; right endocanthian; left exocanthian; and, left endocanthian.

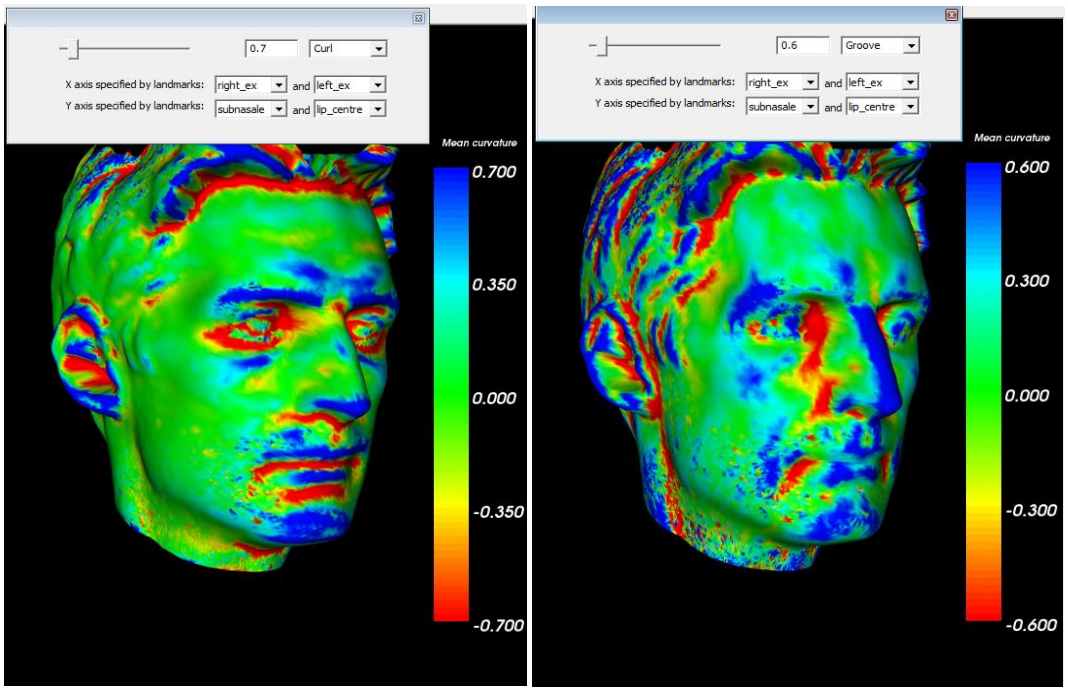


Figure 9.6 Directional curl and groove curvature can be visualised with the addition of axial points: left and right exocanthian; subnasale; and, upper lip centre. This gives a visual indication of philtrum groove and prominence of the philtral pillars.

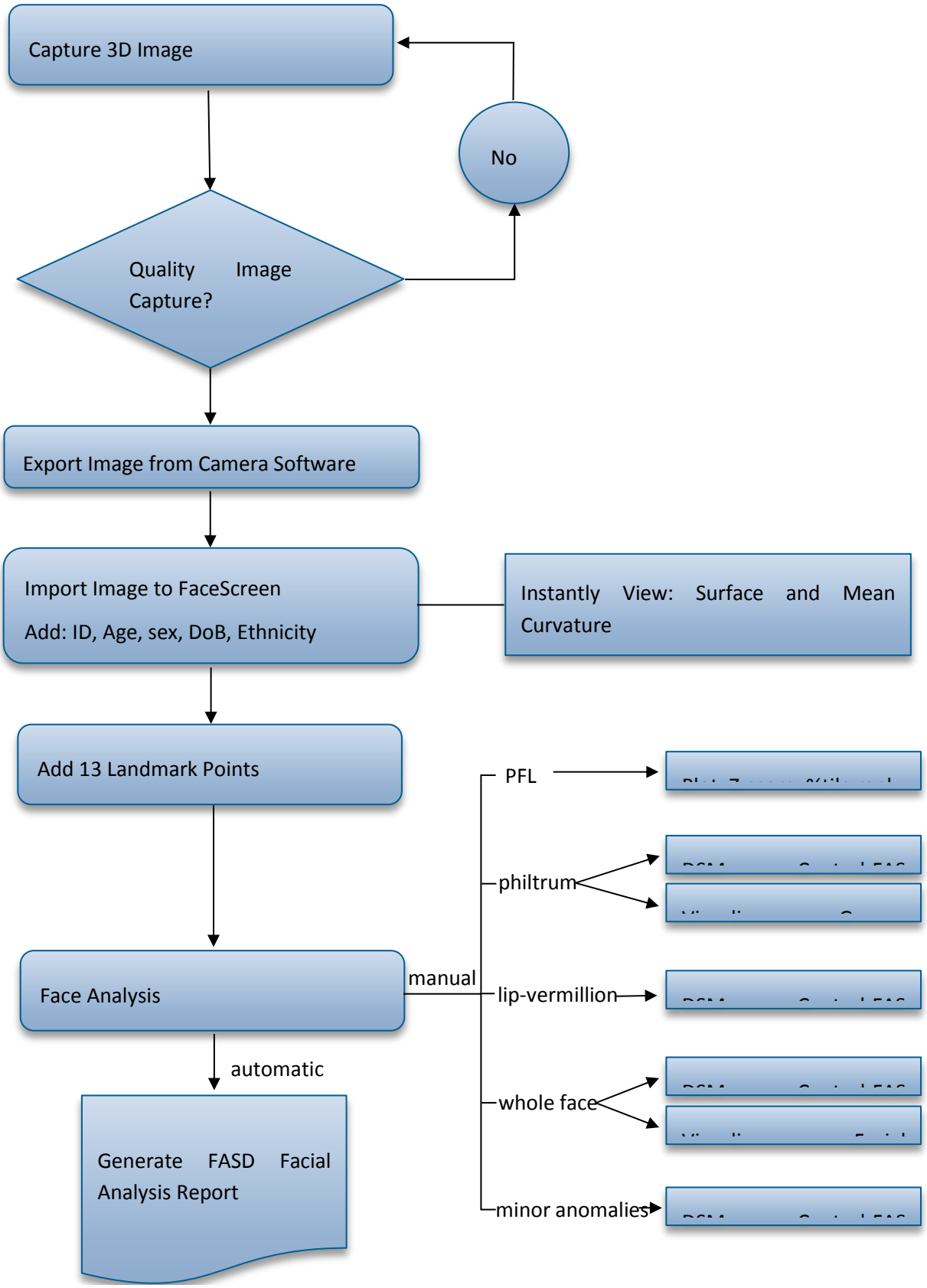
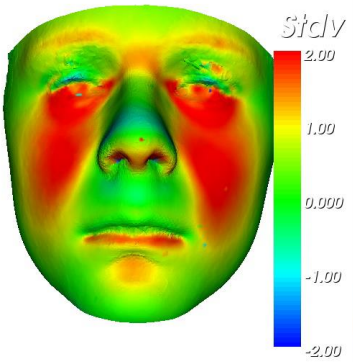


Figure 9.7 User workflow for FaceScreen from image capture to face analysis and report generation.

APPENDIX 7: FACESCREEN FASD REPORT

Report for 3D file: pat_1.obj
Report Date: 21/11/2017

Scan Date	21/11/2017
Date of birth	21/11/2002
Age	15.0
Gender	male
PFL	24.78
PFL Percentile	4.81
PFL Z-Score	-1.66
Ethnic background	Caucasian



PFL:

PFL Measure	24.78mm
Z-Score	-1.67
Percentile	4.81

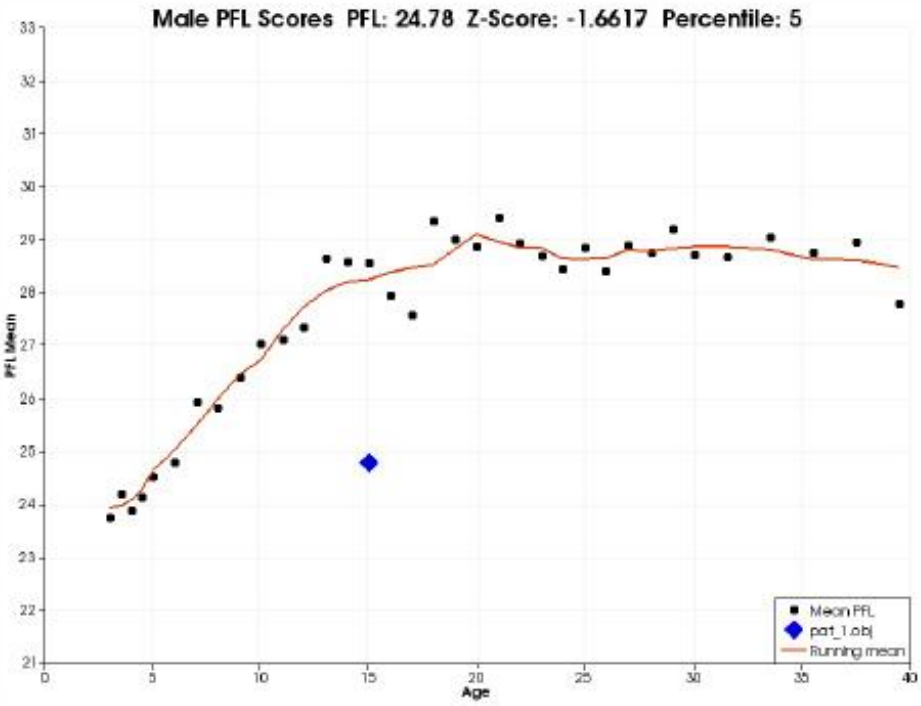
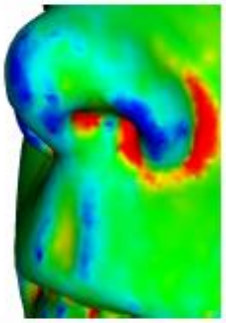


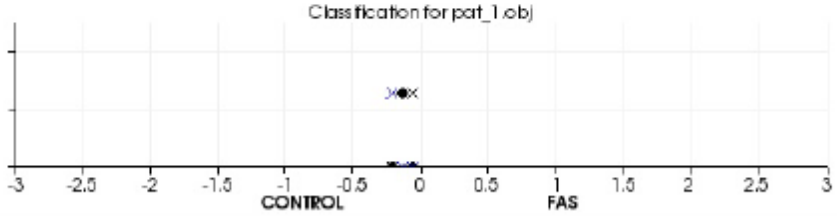
Figure 9.8 FASD 3D facial report page 1. Patient details, portrait and profile images, facial signature followed by PFL analysis.

Philtrum Analysis

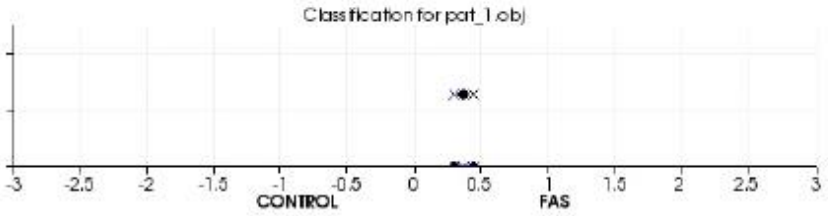
Philtrum Volume	19.43mm2
Philtrum Length	19.26mm
Philtrum Groove Index	1.01



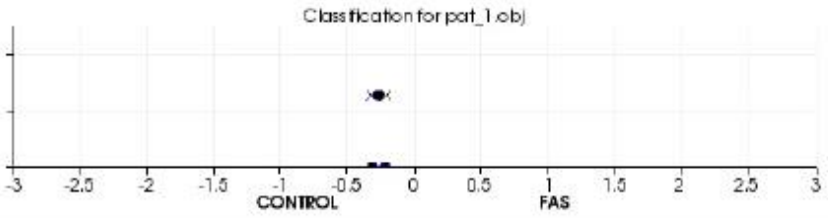
Classification Testing
Philtrum : Value: -0.14, StdError: +/- 0.08



Face : Value: 0.37, StdError: +/- 0.07



Malar : Value: -0.27, StdError: +/- 0.05



Nose : Value: 0.90, StdError: +/- 0.07

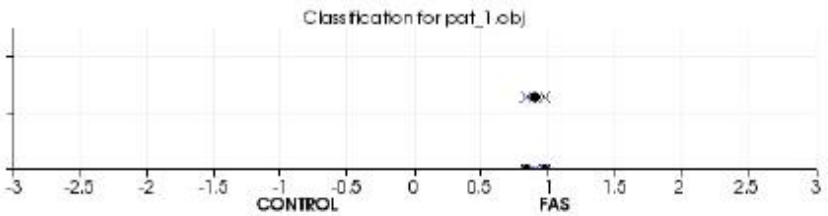


Figure 9.9 FASD 3D facial report page 2 containing philtrum curvature analysis and control-FAS discrimination tests